Table of contents

CUNY 2014 Organizing Committee

Nikole Patson
Shari Speer
Lauren Squires
Rory Turnbull
Laura Wagner
Abby Walker

Review panel

Amit Almor, Sudha Arunachalam, Markus Bader, Giulia Bencini, Moises Betancort, Klinton Bicknell, Allison Blodgett, Julie Boland, Oliver Bott, Dianne Bradley, Dave Braze, Jonathan Brennan, Sarah Brown-Schmidt, Marc Brysbaert, Helen Cairns, Katy Carlson, Manuel Carreiras, Craig Chambers, Franklin Chang, Youngon Choi, Kiel Christianson, Anna Cieslicka, Alexandra Cleland, Charles Clifton, Cynthia Clopper, Derya Cokal, Martin Corley, Seana Coulson, Wind Cowles, Sarah Creel, Marie-Catherine de Marneffe, Dirk den Ouden, Alberto Dominguez, Giuli Dussias, Micha Elsner, Thomas Farmer, Yasmeen Faroqi-Shah, Evelina Fedorenko, Heather Ferguson, Eva Fernández, Fernanda Ferreira, Victor Ferreira, Ruth Filik, Janet Fodor, Stephani Foraker, Julie Franck, Jerid Francom, Lyn Frazier, Cheryl Frenck-Mestre, Susanne Gahl, Alan Garnham, Carlos Gelormini, Edward Gibson, Peter Gordon, Dan Grodner, John Hale, Joy Hanna, Mary Hare, Robert Hartsuiker, Daphna Heller, Barbara Hemforth, Yuki Hirose, Masako Hirotani, Philip Hofmeister, William Horton, Yi Ting Huang, Falk Huettig, E. Matthew Husband, Erika Hussey, Kiwako Ito, Scott Jackson, T. Florian Jaeger, Elsi Kaiser, Yuki Kamide, Andrew Kehler, Frank Keller, Robert Kluender, Pia Knoeferle, Victor Kuperman, Vicky Lai, Ellen Lau, Kerry Ledoux, Richard Lewis, Shane Lindsay, Maryellen MacDonald, Jim Magnuson, Alec Marantz, Andrea Martin, Katherine Messenger, Jelena Mirkovic, Don Mitchell, Nicola Molinaro, Linda Moxey, Shukhan Ng, Padraig O'Seaghdha, Akira Omaki, Anna Papafragou, Nikole Patson, Neal Pearlmutter, Thomas Pechmann, Colin Phillips, Martin Pickering, Maria Pinango, Maria Polinsky, Lucia Pozzan, Ranjani Prabhakaran, Liina Pylkkänen, Pirita Pyykkönen-Klauck, Hugh Rabagliati, Keith Rayner, Hannah Rohde, Douglas Roland, Anne Pier Salverda, Amy Schafer, Christoph Scheepers, William Schuler, Petra Schumacher, Florian Schwarz, Irina Sekerina, Robert Slevc, Ron Smyth, Shari Speer, Lauren Squires, Adrian Staub, Mark Steedman, Linnaea Stockall, Britta Stolterfoht, Patrick Sturt, Benjamin Swets, Michael Tanenhaus, Robert Thornton, Matt Traxler, Annie Tremblay, Danijela Trenkic, John Trueswell, Rory Turnbull, Mieko Ueno, Virginia Valian, Julie Van Dyke, Marten van Schijndel, Shravan Vasishth, Constanze Vorwerg, Matt Wagers, Laura Wagner, Michael Wagner, Abby Walker, Tessa Warren, Thomas Wasow, Duane Watson, Kodi Weatherholtz, Thomas Weskott, Katherine White, Seth Wiener, Ming Xiang, Hiroko Yamashita, ChinLung Yang

Jerrold J. Katz Young Scholar Award

Named in memory of our friend and distinguished colleague, the Jerrold J. Katz Young Scholar Award recognizes the paper or poster presented at the Annual CUNY Conference on Human Sentence Processing that best exhibits the qualities of intellectual rigor, creativity, and independence of thought exemplified in Professor Katz's life and work.
Any first author of a presentation, who is pre-doctoral or up to three years post-PhD, and who is not yet tenured, will be eligible for consideration. The amount of the award is $500.

Previous Recipients

2011 Sol Lago and Wing Yee Chow (University of Maryland, College Park)
Word frequency affects pronouns and antecedents identically: Distributional evidence

2010 Adriana Hanulíková (Max Planck Institute for Psycholinguistics)
When grammatical errors do not matter: An ERP study on the effect of foreign-accent on syntactic processing

2009 Adrian Staub (University of Massachusetts, Amherst)
The timing of garden path effects on eye movements: Structural and lexical factors

2008 Gunnar Jacob (University of Dundee)
An inter-lingual garden-path? L1 interference in L2 syntactic processing

2007 T. Florian Jaeger (University of Rochester) and Neal Snider (Stanford University)
Implicit learning and syntactic persistence: Surprisal and cumulativity

2006 Scott Jackson (University of Arizona)
Prosody and logical scope in English

2005 Sachiko Aoshima (American University)
The source of the bias for longer filler-gap dependencies in Japanese

2004 Andrew Nevins (Massachusetts Institute of Technology)
Syntactic and semantic predictors of tense: An ERP investigation of Hindi

2003 Britta Stolterfoht (Max Planck Institute of Cognitive Neuroscience)
The difference between the processing of implicit prosody and focus structure during reading: Evidence from brain-related potentials

2002 John Hale (Johns Hopkins University)
The information conveyed by words in sentences

Award Fund

To make a contribution to the Jerrold J. Katz Fund, please send a check made out to "CUNY Graduate Center (Sentence Processing Conference)" to the address shown below. It would be helpful if you were to write "Jerrold J. Katz Fund" in the memo line of the check.

Dianne Bradley (Katz Award Fund)
Ph.D. Program in Linguistics
CUNY Graduate Center
365 Fifth Avenue
New York, NY 10016-4309

Special Session on Experimental Pragmatics: Advances in Theory and Method

The dominant strands of research within the field of human sentence processing have focused on the core linguistic areas of syntactic, semantic and lexical processing. Within these areas, a strong and extensive body of work has been created, and the results have led to important insights into human language processing as well as spurred lively ongoing debates about the organization of the cognitive architecture that subserves it. In 2001, the CUNY conference held at the University of Pennsylvania foregrounded an ongoing effort to merge the "language-as-action" and "language-as-product" traditions, an approach that seeks to maintain the latter's focus on specifying the real-time cognitive mechanisms of language processing with the former's focus on language use as it is necessarily intertwined with the ambient world in which it occurs. One area of linguistic inquiry that has shown marked theoretical progress from this effort is pragmatics, the study of how language is produced and understood in a broader communicative context. The scope of context available to speakers and listeners integrates information from variety of sources, including linguistic structure, the visual scene, and interlocutors' world knowledge as well as their mutually sustained beliefs and intentions. Although there is a long and established tradition of investigating the pragmatic aspects of communicative function, researchers in the area had not typically used the kinds of sophisticated experimental methodologies that have become commonplace in other areas of psycholinguistic investigation. However, more recently pragmatics has come into its own within psycholinguistics. The special session on Experimental Pragmatics aims to take stock of recent progress in this burgeoning sub-field and help to set the agenda for future research in this domain.

All presentations marked with ** are part of the special session.

Invited speakers

Sarah Creel is Associate Professor of Cognitive Science at the University of California San Diego. She conducts experimental research on the processing of complex stimuli in speech and music, including spoken word recognition and the structure of memory for acoustic information. Her work has investigated the use of talker information and other contextual cues during language processing and the development of this perceptual capacity in children.

Chris Cummins is a Chancellor's Fellow in the Department of Linguistics and English Language at the University of Edinburgh. Prior to being at Edinburgh, he was a Research Associate in the Faculty of Linguistics and Literary Studies in Bielefeld University. He received his PhD in English and Applied Linguistics from the University of Cambridge. His research is predominantly in experimental semantics and pragmatics, and focuses on the interplay between strategic and automatic processes that motivate speakers' choice of expressions. In his most recent work, he has explored the usefulness of constraint-based models in predicting how form and context interact in the generation of implicatures and the projection of presuppositions. In particular, his work sheds light on how the prior mention of expressions can promote their re-use, and the pragmatic consequences of such re-use.

Judith Degen is a postdoctoral research fellow in the Department of Psychology at Stanford University. She received her PhD in Brain and Cognitive Sciences and Linguistics from the University of Rochester in 2013. Her research investigates pragmatic phenomena in both production and perception. In particular, her recent work on the processing of scalar implicatures has received much attention, due to the use of innovative methodologies to address theoretical questions.

Chris Potts is Associate Professor in the Department of Linguistics and Director of the Center for the Study of Language and Information at Stanford University. He has taught at the North American Summer School in Logic, Language and Information as well as at LSA Summer Institute. His research uses computational and corpus-linguistic methods to investigate topics in formal and computational pragmatics and semantics. Recent research topics include conventional and conversational implicatures, sentiment analysis and expressive content, indexicality, parenthetical expressions and dialogue systems.

Jesse Snedeker is Professor of Psychology at Harvard University. Her research explores the structure of linguistic representation, including semantic and pragmatic processing. She uses diverse experimental and development methods to better understand the relationship between different levels of representation, including studying the acquisition of words among adopted children, using syntactic priming to probe the acquisition and representation of argument structure, and conducting visual world experiments to test children's understanding of implicature.

Gregory Ward is Professor of Linguistics at Northwestern University, where he is also affiliated faculty with the Department of Philosophy and the Program in Gender and Sexuality Studies. His research investigates a variety of issues in discourse and pragmatics, including pragmatic theory, information structure, intonational meaning, and reference/anaphora. A Fellow of the Linguistic Society of America, he has also taught semi-regularly at the LSA Linguistic Institutes, including courses in Experimental Pragmatics. His recent work uses experimental and corpus-linguistic methods to inform pragmatic theory.

Program at-a-glance

All talks will be held in the Wexner Film/Video Theater.

All breaks and the poster sessions will be held in the Mershon Auditorium lobby.

Thursday 3/13

8:00am–9:00am	Continental breakfast and welcome
9:00am–10:45am	Oral session
10:45am–11:15am	Break with coffee
11:15am–1:15pm	Oral session
1:15pm–3:15pm	Poster session I (lunch provided)
3:15pm–5:15pm	Oral session
5:15pm–5:45pm	Break with coffee
5:45pm–7:30pm	Oral session

Friday 3/14

8:00am–9:00am	Continental breakfast
9:00am–10:45am	Oral session
10:45am–11:15am	Break with coffee
11:15am–12:45pm	Oral session
12:45pm–2:45pm	Poster session II (lunch provided)
2:45pm–4:15pm	Oral session
4:15pm–4:45pm	Break with coffee
4:45pm–5:30pm	Oral session
7:30pm–9:30pm	Evening event (11th floor of Thompson Library; pre-registration required)

Saturday 3/15

8:00am–9:00am	Continental breakfast
9:00am–10:45am	Oral session
10:45am–12:45pm	Poster session III (with coffee)
12:45pm–1:45pm	Lunch (on your own)
1:45pm–4:00pm	Oral session

Full program

Presentations marked with ** are part of the Special Session on Experimental Pragmatics.

Thursday

8:00am–08:45am	Continental breakfast
8:45am–9:00am	Welcome
9:00am–9:45am	**Characterizing expressive and social meaning with large corpora *Chris Potts*
9:45am–10:15am	The mind leads the eyes: ungrammaticality detection from two words back in reading *Klinton Bicknell and Roger Levy*
10:15am–10:45am	The relationship between regressive saccades and the P600 effect: Evidence from concurrent eye movement and EEG recordings *Paul Metzner, Titus von der Malsburg, Shravan Vasishth and Frank Rösler*
10:45am–11:15am	Break with coffee
11:15am–11:45am	Unforced revision in the processing of relative clause association ambiguity in Japanese *Toshiyuki Yamada, Manabu Arai and Yuki Hirose*
11:45am–12:15pm	Ambiguity affects Mandarin relative clause processing *Yaling Hsiao, Jinman Li and Maryellen MacDonald*
12:15pm–12:45pm	Underspecification in RC attachment: A speed-accuracy tradeoff analysis *Pavel Logačev and Shravan Vasishth*
12:45pm–1:15pm	Linear order effects in agreement: Evidence from English wh-questions *Brian Dillon, Joshua Levy, Adrian Staub and Charles Clifton*
1:15pm–3:15pm	Poster Session I (lunch provided)
3:15pm–3:45pm	The influence of verb-specific featural restrictions, word associations, and production-based mechanisms on language-mediated anticipatory eye movements *Florian Hintz, Antje S. Meyer and Falk Huettig*
3:45pm–4:15pm	Spillover frequency effects in a sequential sampling model of reading *Michael Shvartsman, Richard Lewis and Satinder Singh*
4:15pm–4:45pm	Auditory perceptual simulation during silent reading *Peiyun Zhou and Kiel Christianson*
4:45pm–5:15pm	Gradience and individual differences in processing syntactic, semantic and thematic information: ERP evidence *Darren Tanner and Janet G. van Hell*
5:15pm–5:45pm	Break with coffee

(Thursday program continued overleaf)

Thursday (continued)

5:45pm–6:15pm	**The contribution of form and meaning to the processing of careful and casual speech. *Kevin McGowan, Meghan Sumner, Annette D'Onofrio and Teresa Pratt*
6:15pm–6:45pm	**Expectation-adaptation in the incremental interpretation of contrastive prosody *Chigusa Kurumada, Meredith Brown, Sarah Bibyk, Daniel Pontillo and Michael Tanenhaus*
6:45pm–7:30pm	**Alternatives in pragmatic inference *Judith Degen*

Friday

8:00am–9:00am	Continental breakfast
9:00am–9:45am	**The case of the inconsistent implicature *Chris Cummins*
9:45am–10:15am	**"Not all" comes for free *Francesca Foppolo and Marco Marelli*
10:15am–10:45am	**Some implicatures take their time. An ERP study on scalar implicatures with 'sentence- picture vs. picture-sentence' verification task. *Daniele Panizza and Edgar Onea*
10:45am–11:15am	Break with coffee
11:15am–11:45am	The local contrast expectation in let alone coordination *Jesse Harris and Katy Carlson*
11:45am–12:15pm	Language affects quantity judgments in bilingual Yudja speakers *Suzi Lima*
12:15pm–12:45pm	Time heals semantic illusions, but not syntactic illusions *Dan Parker, Alan Du and Colin Phillips*
12:45pm–2:45pm	Poster Session II (lunch provided)
2:45pm–3:15pm	The P600 and the n-back task: Evidence that domain general conflict resolution ability underpins the resolution of garden-paths *Polly L. O'Rourke and Gregory J. H. Colflesh*
3:15pm–3:45pm	Predictability and filler-gap ordering in dependency formation: An MEG study *Kimberly Leiken and Liina Pylkkänen*
3:45pm–4:15pm	Facilitatory semantic interference in the processing of long distance dependencies *Ian Cunnings and Patrick Sturt*
4:15pm–4:45pm	Break with coffee
4:45pm–5:30pm	**Demonstrative equatives and the conveyance of speaker (un)certainty *Gregory Ward*

Saturday

8:00am–9:00am	Continental breakfast
9:00am–9:45am	**Hearing a who: Preschoolers and adults process language talker-contingently *Sarah Creel*
9:45am–10:15am	**Privileged vs. shared knowledge about object identity in real-time referential processing *Mindaugas Mozuraitis, Craig G. Chmabers and Meredyth Daneman*
10:15am–10:45am	**What do you know? ERP evidence for immediate use of common ground during online reference resolution *Les Sikos, Sam Tomlinson, Lauren Sanchez and Dan Grodner*
10:45am–12:45pm	Poster Session III (with coffee)
12:45pm–1:45pm	Lunch (on your own) & Eyelink workshop
1:45pm–2:15pm	A cross-linguistic verb-final bias in gesturing paradigms *Richard Futrell, Tina Hickey, Aldrin Lee, Eunice Lim, Elena Luchkina and Edward Gibson*
2:15pm–2:45pm	Processing of pitch prominence in Williams syndrome *Kiwako Ito, Marilee Martens and Erin McKenna*
2:45pm–3:15pm	The processing of prosodic focus in French. *Jui Namjoshi and Annie Tremblay*
3:15pm–4:00pm	**Scalar implicature: a whirlwind tour with stops in processing, development and disorder *Jesse Snedeker*

Poster sessions

POSTER SESSION I (Thursday, 1:15pm-3:15pm)	
1001	Abstract structure is active during comprehension of collocations Erin Conwell and Jesse Snedeker
1002	Encoding and retrieval interference in dependency resolution Jens Roeser, Lena Jäger, Lena Benz and Shravan Vasishth
1003	Exploring socioeconomic differences in syntactic development through processing Yi Ting Huang, Kathryn Leech and Meredith Rowe
1004	Homophone disambiguation & the social identity of the speaker Julie Boland and Maryam Seifeldin
1005	How sentence processing benefits from the inflectional richness of a language Sofia Manika
1006	Experience with dialectal variants modulates online syntactic comprehension Scott Fraundorf and T. Florian Jaeger
1007	Language processing shapes language change: Redundant cues are less likely to be maintained in language Maryia Fedzechkina, Elissa L. Newport and T. Florian Jaeger
1008	Listeners maintain and rationally update uncertainty about prior words in spoken comprehension Klinton Bicknell, Michael Tanenhaus and T. Florian Jaeger
1009	Learning language from the environment depends on the fitness of both the learner and the environment Shiri Lev-Ari and Sharon Peperkamp
1010	Effects of interlocutor feedback on speaker phonetic production in a simulated-communication task Esteban Buz, T. Florian Jaeger and Michael Tanenhaus
1011	Error detection in native and non-native speakers provides evidence for a noisy channel model of sentence processing Robin Hill and Frank Keller
1012	Partial separation of syntactic representations in closely related varieties. Evidence from Swiss German and Standard German Constanze Vorwerg, Janine Lüthi and Manuel Straessle
1013	The interaction of images and text during comprehension of garden-path sentences Cassie Palmer and Kiel Christianson
1014	Bootstrapping into filler-gap: an acquisition story Marten van Schijndel and Micha Elsner
1015	Online comprehension of object wh-questions: eye-tracking evidence against syntactic gap filling Olga Dragoy, Maria Ivanova, Anna Laurinavichyute, Anastasia Ulicheva, Svetlana Kuptsova and Lidia Petrova

1016	Processing filler-gap dependencies in L2: Evidence for the use of subcategorization information Chie Nakamura, Manabu Arai, Yasunari Harada and Yuki Hirose
1017	Interference effects of pre-verbal NPs on sentence processing in Japanese Hajime Ono, Miki Obata and Noriaki Yusa
1018	The effect of L1 syntax on L2 sentence processing: A self-paced reading study with L2 learners of Chinese Yun Yao
1019	Long-distance attraction effects in SVA processing Nathan Eversole and Jeffrey Witzel
1020	An effect of verb repetition in the production of head-final passive construction Lan Jin, Ying Deng, Manabu Arai and Yuki Hirose
1021	It takes time to prime: Semantic priming in the ocular lexical decision task Renske Hoedemaker and Peter C. Gordon
1022	Computing the structure of questions: Evidence from online sentence processing Hadas Kotek and Martin Hackl
1023	Online processing of English Which-questions by children and adults: a visual world study Carla Contemori and Theodoros Marinis
1024	A bayesian meta-analysis of the Chinese relative clause question Shravan Vasishth
1025	Non-adjacent lexical dependencies in an artificial language prime relative clause attachment biases Hao Wang, Mythili Menon and Elsi Kaiser
1026	The processing and neural basis of argument structure composition through eye-tracking, focal brain-lesion and fMRI Sara Sanchez-Alonso, Amy Ly, David Braze, Cheryl Lacadie, Todd Constable and Maria Pinango
1027	Turkish-speaking children use verbal morphemes to predict argument structure Duygu Özge, Jesse Snedeker, Deniz Özkan-Göktürk, Berna Uzundağ and Aylin Küntay
1028	Two distinct attraction profiles in comprehending Russian gender agreement Natalia Slioussar, Anton Malko and Colin Phillips
1029	No lexical boost: verb-based information does not facilitate prediction over and above event-based knowledge in the visual world Evelyn Milburn, Tessa Warren and Michael Walsh Dickey
1030	Event representations in collective and distributive readings: an on-line study Garrett Smith
1031	Comprehenders mentally represent some aspects of plural entities Nikole Patson
1032	Misinterpretation in agreement and agreement attraction Nikole Patson and E. Matthew Husband
1033	Shifting viewpoints and discourse economy Jesse Harris
1034	Gestures modulate access to referent representations Stephani Foraker

1035	Do the pitch accents or the phrasal accents determine an alternative question? Samantha Heidenreich, Shari Speer and Carl Pollard
1036	Predicting L2 lexical pitch accent perception: the role of top-down knowledge in L2 Japanese Seth Goss and Katsuo Tamaoka
1037	Priming of implicit prosody and individual differences Sun-Ah Jun and Jason Bishop
1038	Sentence completion to the beat: Effects of implicit prosodic rhythm in English and German Mara Breen and Gerrit Kentner
1039	From single words to full sentences Zeshu Shao and Agnieszka Konopka
1040	Coordination of eye movements and speech in the RAN task predicts sentence reading Peter C. Gordon and Renske S. Hoedemaker
1041	Integrating new vs. revising previous input: Local coherences vs. garden paths Kiel Christianson, Steven G. Luke and Kacey Wochna
1042	Skipping syntactically illegal “the”-previews - the role of predictability Matthew Abbott, Bernhard Angele, Danbi Ahn and Keith Rayner
1043	Predicting individual differences in underspecification: An integrated model of good-enough processing Felix Engelmann and Shravan Vasishth
1044	Recognizing words during sentence processing: Behavioral and neural measures Peter C. Gordon, Giulia C. Pancani, Matthew W. Lowder, Renske S. Hoedemaker and Mariah Moore
1045	Selective priority for structure in memory retrieval Dan Parker and Colin Phillips
1049	** Aim low: Speakers design utterances for the most naive addressee Si On Yoon and Sarah Brown-Schmidt
1050	** Definiteness and givenness affect pronoun resolution: Evidence from eye fixations Klaus von Heusinger, Sofiana Chiriacescu, Andreas Brocher and Tim Graf
1051	** Differences in the prosodic encoding of information structure in closely-related languages Arunima Choudhury and Elsi Kaiser
1052	** Dissociating neural effects of predictability and incongruity in adjective-noun phrases Ellen Lau and Anna Namyst
1053	** Effects of context on the processing of adversative and comparative constructions Grégoire Winterstein, Emilia Ellsiepen, Jacques Jayez and Barbara Hemforth
1054	** Elaboration of concepts facilitates their retrieval in sentence processing Melissa Troyer, Philip Hofmeister and Marta Kutas
1055	** Hippocampal contributions to discourse processing: Findings from Amnesia Sarah Brown-Schmidt, Melissa Duff and Jake Kurczek
1056	** Impact of focus alternatives on memory for focus referents Wind Cowles and Michelle Perdomo

1057	** Is EVERY child different until they turn 15?: Not so quick development of universal quantifier comprehension Kiwako Ito, Utako Minai and Adam Royer
1058	** Pragmatic processing costs reflect linking to context, not enrichment Shevaun Lewis and Colin Phillips
1059	** Prosodic encoding of information structure depends on frequency and probability Iris Chuoying Ouyang and Elsi Kaiser
1060	** Restrictive vs. non-restrictive composition: An MEG Study Timothy Leffel, Miriam Lauter, Masha Westerlund and Liina Pylkkänen
1061	** Soft and hard presupposition triggers are fast in online processing Florian Schwarz
1062	** The role of working memory in the online realization of scalar inferences Stephen Politzer-Ahles, Robert Fiorentino, Jeffery Durbin and Lillian K. Li
1063	** Word order typology and discourse expectation Eunkyung Yi and Jean-Pierre Koenig
1064	** Young children's comprehension of negation Tracy Brookhyser, Roman Feiman and Jesse Snedeker
1065	** Meaning what you say, or saying it the way you mean it: Suprasegmental cues to non-literal meaning Georgia Simon and Kristen Syrett

POSTER SESSION II (Friday, 12:45pm-2:45pm)	
2001	Delaying verb production changes what matters in subject-verb agreement Arielle Allentoff and Heidi Lorimor
2002	Semantic relatedness and semantic integration in subject-verb agreement errors Darrell Penta and Neal Pearlmutter
2003	Yo pienso, tu piensas: Crosslinguistic agreement effects in comprehension Sol Lago, Diego Shalom, Mariano Sigman, Ellen Lau and Colin Phillips
2004	Impact of phonological distance on lexical ambiguity resolution in people with and without aphasia Maria Ivanova, Anna Laurinavichyute, Olga Dragoy, Anastasia Ulicheva and Svetlana Kuptsova
2005	Antecedent contained deletions in native and non-native sentence processing Oliver Boxell, Claudia Felser and Ian Cunnings
2006	Gender agreement errors in learner English: Evidence from Production and Comprehension Lucia Pozzan, Inés Antón-Méndez and John Trueswell
2007	Processing at the semantic and syntactic interface in learners of Spanish: evidence from ERPs Carla Contemori, Patricia Roman and Giuli Dussias
2008	Complement coercion as aspectuality: Evidence from SPR and fMRI Yao-Ying Lai, Cheryl Lacadie, Todd Constable, Ashwini Deo and Maria Piñango
2009	Timing of lexical activation in determiner–adjective–noun phrase production Amy DiBattista and Neal Pearlmutter
2010	Effects of animacy and noun-phrase relatedness on the processing of complex sentences Matthew Lowder and Peter C. Gordon
2011	Effects of animacy and semantic relatedness during sentence processing: An ERP study Mariana Vega-Mendoza, Martin Pickering and Mante Nieuwland
2012	A common neural basis for syntactic and non-syntactic conflict-control Nina Hsu, Susanne Jaeggi and Jared Novick
2013	Effects of experience and expectations on adaptation to dialect variation in noise Kodi Weatherholtz, Abby Walker, Shannon Melvin, Adam Royer and Cynthia Clopper
2014	Locality in filler-gap dependencies: Evidence from extraposition Brian Dillon
2015	Semantics overrides syntax in the processing of gap-filler dependencies in Chinese Shukhan Ng and Nicole Wicha
2016	Structural and non-structural locality effects in Bangla filler-gap dependencies Dustin Chacón, Mashrur Imtiaz, Shirsho Dasgupta, Sikder Monoare Murshed, Mina Dan and Colin Phillips
2017	Online processing of parasitic gaps: Evidence from eye-tracking Cassandra Chapman
2018	Understanding ambiguous idioms in context: Clearing the air on compositional and noncompositional views Nyssa Z. Bulkes and Christopher M. Grindrod

2019	Abstract agreement: Children's sensitivity to subject-verb agreement in comprehension does not require knowledge of specific lexical co-occurrences. Cynthia Lukyanenko and Cynthia Fisher
2020	Mapping the kindergarten-path: Cognitive predictors of child sentence processing Kristina Woodard, Lucia Pozzan and John C. Trueswell
2021	Sensitivity to online encoding and retrieval interference in younger and older adults Dave Kush, Clinton L. Johns and Julie Van Dyke
2022	Skewed lexical distributions facilitate recursion learning in an artificial grammar task Pyeong Whan Cho, Garrett Smith and Whitney Tabor
2023	The effect of syntactic category on advance planning in sentence production Shota Momma, Robert Slevc and Colin Phillips
2024	Verb-initial structures in Arabic: Qualitative ERP differences between singular and plural subjects. R. Muralikrishnan and Ali Idrissi
2025	Intrusion effects on NPI licensing in Turkish: Does the parser ignore the grammar? Aydogan Yanilmaz and John Drury
2026	Priming competes with syntactic anticipation, both within and across languages: Evidence from the Visual World Paradigm Francesca Filiaci, Thomas Bak, Frank Keller and Antonella Sorace
2027	** Priming enriched meanings Lewis Bott and Emmanuel Chemla
2028	There's more to a sentence than its words: Repetition priming in sentences and lists Giulia C. Pancani, Peter C. Gordon and Joseph Hopfinger
2029	Acceptability, statistical distribution, and non-usage-based mechanisms in processing null pronouns, overt pronouns, and repeated names in Brazilian Portuguese Jefferson de Carvalho Maia, Maria Luiza Cunha Lima, Amit Almor and Carlos Gelormini Lezama
2030	Reduction in duration depends on articulation, not simply facilitated processing Cassandra L. Jacobs, Duane Watson and Gary S. Dell
2031	Eye-movements during reading reflect subsequent sentence memory: evidence from subject and object relative clauses Brennan Payne, Ariel James, Elizabeth Stine-Morrow and Duane Watson
2032	Processing resumptives in Mandarin relative clauses: An eye-tracking study Li-Hsin Ning, Kiel Christianson and Chien-Jer Charles Lin
2033	Relative clause processing and competing pressures in an agreement-rich language Manuel F. Borja, Sandra Chung and Matt Wagers
2034	Resumptive pronouns and structural complexity in Cantonese relative clause production Elaine Francis, Charles Lam, Carol Chun Zheng, John Hitz and Stephen Matthews
2035	Resumptive pronouns salvage island violations in forced-choice tasks Lauren Ackerman, Michael Frazier and Masaya Yoshida
2036	Contrast-driven phonetic variation in English and Russian Clara Cohen

2037	Auditory perceptual simulation of child-directed speech in silent reading Zhiying Qian, Mallory Stites and Kiel Christianson
2038	Listeners update probabilistic percepts of (a) function word(s) several syllables downstream Meredith Brown, Laura C. Dilley and Michael K. Tanenhaus
2039	Great expectation follows big surprise: Direct evidence for syntactic adaptation from eye-movements Manabu Arai and T. Florian Jaeger
2040	Contributions of hippocampal-dependent declarative memory to on-line processing of global syntactic ambiguity Sarah Brown-Schmidt, Zhenghan Qi and Melissa Duff
2041	Morphology aids syntax in noisy sentence processing Gaurav Kharkwal and Karin Stromswold
2042	Working-memory capacity modulates antilocality effects in syntactic dependencies Bruno Nicenboim, Shravan Vasishth, Reinhold Kliegl, Carolina Gattei and Mariano Sigman
2043	Efficient communication forward and backward Richard Futrell, Kyle Mahowald, Steve Piantadosi and Edward Gibson
2044	Partial use of available information in the early stages of verb prediction Wing Yee Chow, Glynis MacMillan, Shefali Shah, Ilia Kurenkov, Ellen Lau and Colin Phillips
2045	Beyond syntax: Effects of verb semantics and perspective taking on Chinese reflexives Xiao He and Elsi Kaiser
2049	**2.5-year-olds consider presuppositions: An eye-tracking study on TOO and AGAIN Frauke Berger and Nausicaa Pouscoulous
2050	** Before and after, and processing presuppositions in discourse Ming Xiang, Emily Hanink and Genna Vegh
2051	** Cognitive resources underlying scalar implicature: A subclinical study E. Matthew Husband and Nicholas Angelides
2052	** Indirect scalar implicatures are neither scalar implicatures nor presuppositions (or both) Cory Bill, Jacopo Romoli, Florian Schwarz and Stephen Crain
2053	** Neural correlates of realizing scalar inferences: An MEG study Stephen Politzer-Ahles
2054	** Processing scalars Mandy Simons and Tessa Warren
2055	** Prosody affects scalar implicature generation Marie-Catherine de Marneffe and Judith Tonhauser
2056	** Pragmatic inferencing across scales: Linguistic and extra-linguistic effects Yi Ting Huang
2057	** Exhaustivity and polarity-mismatch Aron Hirsch
2058	** Focus preferences for focus-sensitive particles (and why) Jesse Harris and Katy Carlson
2059	** Pragmatic fit and the processing of Korean honorifics Hong Mo Kang, Jean-Pierre Koenig and Gail Mauner

2060	Processing asymmetries between Subject-Only and VP-Only Martin Hackl, Erin Olson and Ayaka Sugawara
2061	** Rapid inference of intended information transfer rate in conversation Mark Myslín
2062	** Recognizing what-is-said versus what-is-implicated William Horton, Chris Schmader and Gregory Ward
2063	** The cognitive and neural basis of pragmatic processing: A case study of jokes Evelina Fedorenko, Jeanne Gallée and Zuzanna Balewski
2064	** The cost of pragmatic inference in the production of referring expressions Peter Baumann
2065	** The effect of inferred explanations in a Bayesian theory of pronominal reference Andrew Kehler and Hannah Rohde

POSTER SESSION III (Saturday, 10:45am-12:45pm)	
3001	Language processing skill is not a unitary construct: Infants' vocabulary knowledge drives lexical and sentence processing Arielle Borovsky, Erica Ellis, Julia Evans and Jeffrey Elman
3002	Text exposure may affect relative clause use in children and adults Jessica Montag and Maryellen MacDonald
3003	Bad, or just less good? ERPs of processing Arabic agreement violations for plural subjects. Ali Idrissi and R. Muralikrishnan
3004	Processing of grammatical agreement across clause boundaries Matthew Lowder and Peter C. Gordon
3005	Surface form effects in agreement attraction and similar phenomena Natalia Slioussar and Natalia Cherepovskaia
3006	Studying anaphoric dependencies using co-registration of eye movements and ERPs Titus von der Malsburg, Paul Metzner, Shravan Vasishth and Frank Rösler
3007	Social category effects on the reconceptualization of referents in dialogue Chris Schmader and William Horton
3008	Speaking in context: discourse shapes incremental preparation of simple sentences Agnieszka Konopka
3009	Memory for what was said in conversation: Speakers and Listeners differ Si On Yoon and Sarah Brown-Schmidt
3010	Continuous acoustic information trickles up; discourse information trickles down Sarah Brown-Schmidt and Joseph Toscano
3011	** The influence of partner-specific memory associations on picture naming: A failure to replicate Horton (2007) Sarah Brown-Schmidt and William S. Horton
3012	Representing multiple instantiations of an object: Effects of visual and linguistic context on real-time event processing Xin Kang, Gitte Joergensen and Gerry Altmann
3013	Verbs drive the bus: An ERP study on the role of verb bias and plausibility information in the resolution of DO/SC ambiguity in English monolinguals and Spanish-English bilinguals. Patricia E. Román, Nicholas R. Ray, Carla Contemori, Edith Kaan and Paola E. Dussias
~~3014~~	~~** Presuppositions in the scope of quantifying expression: Eye tracking data~~ ~~Sonja Tiemann~~ (paper withdrawn)
3015	Animacy and the active construction of filler-gap dependencies in relative clauses Emily Pendleton and Matt Wagers
3016	Grammatically-guided resolution of filler-gap dependencies: An investigation of Chinese multiple dependencies Jie Liu, Robert Fiorentino and Alison Gabriele
3017	Increased linguistic focus leads to increased reading times Matthew Lowder and Peter C. Gordon
3018	Intrinsic differences in the processing of singular and plural pronouns Mahayana C. Godoy and Cíntia Antão

3019	The Middle matters the most: The effect of phonological similarity on referring forms Hossein Karimi and Fernanda Ferreira
3020	Noun-phrase-internal structural priming in a picture-description task Amy DiBattista and Neal Pearlmutter
3021	Decoding the semantics of words in sentences from neural activity Alona Fyshe, Erika Laing, Nicole Rafidi, Kai-Min Chang and Tom Mitchell
3022	Eye movements reflect the cognitive reality—and costs—of event structure during reading Benjamin Swets and Christopher Kurby
3023	Implicit agents in short passives and remote control of reason clauses Michael McCourt, Aleksandra Fazlipour, Ellen Lau and Alexander Williams
3024	Unconscious lexical semantic activation of form neighbors Dane Bell, Kenneth Forster and Thomas Bever
3025	A speed-accuracy tradeoff study of poor readers' memory mechanisms Clinton L. Johns, Julie Van Dyke and Dave Kush
3026	Effects of integration in eye tracking William Schuler and Marten van Schijndel
3027	Eye-tracking and event-related potential effects of transposed letters during compound word comprehension: Do morphemes matter? Mallory Stites, Kara D. Federmeier and Kiel Christianson
3028	On-line sentence reading in people with aphasia: Evidence from eye tracking Gayle DeDe and Jessica Knilans Flax
3029	Why do interference effects arise in sentence processing? A sampling theory of memory as optimal discrimination among noisy traces Richard Lewis, Michael Shvartsman and Satinder Singh
3030	OV-VO yield imperfect mirror-images: On the impact of length on sentence word order Idoia Ros, Mikel Santesteban, Kumiko Fukumura and Itziar Laka
3031	A multiple argument overlap boost in Japanese structural priming Hiroko Yamashita and Franklin Chang
3032	Cross-linguistic differences in processing double-embedded relative clauses: working-memory constraints or language statistics? Stefan L. Frank, Shravan Vasishth and Thijs Trompenaars
3033	Effects of repeated exposure on sentence processing time: Object relatives and "early closure" ambiguities Matt Traxler and Kristen M. Tooley
3034	Structural forgetting in processing complex Mandarin relative clauses Yuxia Wang, Francois Rigalleau and Manuel Gimenes
3035	Anti-locality preference in the processing of Japanese reflexive binding Akira Omaki, Brian Dillon, Takuya Kubo, Manami Sato and Hiromu Sakai
3036	Local coherence effects and crossing reflexive- and wh-dependencies Michael Frazier, Peter Baumann, Lauren Ackerman, David Potter and Masaya Yoshida
3037	Locality and anti-locality effects in the processing of expected and unexpected Inputs: A study of NPI dependencies in Japanese Kentaro Nakatani

3038	Effects of the second language on syntactic processing in the first language Giuli Dussias, Lauren Perrotti, Matt Brown and Luis Morales
3039	A chunk of coffee: an event-related brain potential (ERP) study on the processing of Mandarin classifiers Zhiying Qian and Susan Garnsey
3040	Attuning to cohesion: English count-syntax, the Mandarin general classifier ge, and wholeness Jessica Harding and Charles Lin
3041	Implicit prosody primes appropriately intonated auditory probes Anouschka Foltz, Cody Linne and Shari Speer
3042	Prosodic and syntactic effects in gapping interpretation Michael Frazier and Katy Carlson
3043	When phonological systems clash: L1 phonotactics vs. L2 assimilation David Li and Elsi Kaiser
3044	Capturing psycholinguistic processing effects using Amazon Mechanical Turk Kelly Enochson and Jennifer Culbertson
3045	Pupillometry – the Index of Cognitive Activity as a measure of linguistic processing difficulty Vera Demberg and Asad Sayeed
3046	Relativized Minimality: a systematic investigation on intervention effects Sandra Villata, Luigi Rizzi, Akira Omaki and Julie Franck
3047	ERP indices of predictive processing Edith Kaan, Evan Carlisle, George Collins, Nicholas Feroce and Amalia Reyes
3049	** Effects of morphological and prosodic focus cues on topic maintenance in Korean Kitaek Kim, Theres Grüter and Amy Schafer
3050	** Determining necessary and possible values: An online study of modals and superlative modifiers Yaron McNabb and Doris Penka
3051	** Generating implicatures from English NPs of the form a/an X: Generalized or local? Li-Hsin Ning and Marina Terkourafi
3052	** Influence of visual complexity on referring expression generation Hannah Rohde, Alasdair Clarke and Micha Elsner
3053	** Lexical entrainment in deceptive interaction Alessia Tosi, Holly Branigan, Adam Moore and Martin Pickering
3054	** Mechanisms of prosody production: Differences between children with and without ASD. Jennifer E. Arnold, Elise C. Rosa, Mark Klinger, Patrick Powell, Alison Meyer
3055	** On questions and speaker ignorance Xiaobei Zheng and Richard Breheny
3056	** Pragmatic influences on the processing of ACD relative clauses Edward Gibson, Polly Jacobson, Peter Graff, Kyle Mahowald, Evelina Fedorenko and Steven Piantadosi
3057	** Processing canonical and non-canonical sentences in Turkish within a context Barış Kahraman and Yuki Hirose

3058	** What you believe is what you get: The role of individual beliefs in sentence comprehension Heeju Hwang
3059	** Influencing persistence of meanings in entailment-obstructing grammatical environments E. Allyn Smith and Kathleen Currie Hall
3060	** Reference as a side effect of discourse expectations: The case of focus-sensitive particles Israel de la Fuente and Barbara Hemforth
3061	** Pragmatic narrowing in reference resolution: Domain restriction & perspective taking Florian Schwarz and Dan Grodner

Paper Abstracts

Characterizing expressive and social meaning with large corpora

Christopher Potts (Stanford University)
cgpotts@stanford.edu

In this talk, I'll sing the praises of large corpora as tools for understanding lexical items and constructions that convey meanings we might variously label as social, expressive, attitudinal, and emotional. Such meanings are highly context dependent and challenging to paraphrase. Large and diverse corpora are our best chance at seeing and capturing the full range of usage patterns.

I'll focus in particular on analyses involving Web-derived corpora with naturalistic annotations: author-assigned ratings, reader-assigned feedback buttons, and so forth. These annotations provide glimpses of speaker/author intentions and listener/reader construals. To illustrate, I'll review a case study involving affective uses of demonstratives (Acton & Potts 2014), seeking to capture the complex pragmatic effects of these uses and relating those effects to broader social outcomes.

I'll also seize this opportunity to reflect on the relationship between corpus work and psycholinguistic work (de Marneffe & Potts 2014). In the case of the expressive and social meanings under consideration here, the corpus evidence is abundant, diverse, and naturalistic, which is essential for getting a full and unbiased look at the phenomena, but it's also noisy and purely observational. By comparison, evidence from psycholinguistic experiments is likely to be relatively clear and controlled, but produced in circumstances that might deeply affect the phenomena under investigation. Thus, the two modes of investigation seem likely to provide complementary insights into these challenging but vital areas of linguistic meaning.

References

Acton, Eric K. & Christopher Potts. 2013. That straight talk: Sarah Palin and the sociolinguistics of demonstratives. To appear in *Journal of Sociolinguistics*.

de Marneffe, Marie-Catherine & Christopher Potts. 2014. Developing linguistic theories using annotated corpora. To appear in Nancy Ide & James Pustejovsky, eds., *The Handbook of Linguistic Annotation*. Berlin: Springer.

The mind leads the eyes: Ungrammaticality detection from two words back in reading

Klinton Bicknell (Univ. of Rochester, Northwestern University) & Roger Levy (UC San Diego)
kbicknell@northwestern.edu

A major theoretical result and methodological cornerstone of sentence processing is the close connection between the progression of the eyes and the mind through visual input of all sorts. In reading, eye movement measures on a word (word *n*) are generally taken to reflect the processing of word *n* in its preceding context, or sometimes spillover from preceding material. Under some circumstances, readers also perform some processing of the next word (*n*+1) while fixating word *n*, and recent evidence suggests low-level processing of future word *n*+2 (e.g., Angele & Rayner, 2011). It is unknown how such partial processing interacts with parsing, as there is no clear evidence that readers are sensitive to the grammatical fit of future words *n*+1 or *n*+2. Here, we present evidence that comprehenders can detect the ungrammaticality of a word that is two words past the currently fixated word. This is to our knowledge the first clear demonstration of grammatical processing of future words.

Materials. We constructed stimuli with three conditions (see example). In the Ungrammatical condition, a verb (*skimmed*) two-words downstream from the target word *n* (*class*) rendered the incremental parse ungrammatical. In the Grammatical, Same-Context condition, the sentence is identical up through word *n*+1, but now word *n*+2 is a grammatical continuation. The Grammatical, Same-*n*+2 condition has a word insertion or replacement prior to word *n*-1 that renders grammatical a verb at word *n*+2 (as in the ungrammatical condition). Words *n*-1, n, and *n*+1 were identical across all conditions. Word *n*+1 was short (3–4 letters).

Methods. We tracked the eye movements of 50 participants reading 22 items in these three conditions with 26 fillers. By comparing measures of processing difficulty on word *n* across the Ungrammatical and Same-Context conditions, we obtain a measure of the rate at which the ungrammatical word *n*+2 is detected. Because this comparison may be confounded by the mere presence of a long word (*skimmed*) to the right of fixation, we also compare Ungrammatical to Same-*n*+2. If there are more signs of processing difficulty on word *n* for Ungrammatical than either Same-*n*+2 or Same-Context, this represents evidence that comprehenders are sensitive to the grammatical fit of word *n*+2 when fixating word *n*. Statistical tests are via mixed-effects regression with maximal random effects structure. We removed outliers more than 3 standard deviations from participant means and trials with blinks or fixations longer than 1 second.

Results. Gaze durations (the time from first fixating a word until moving off) and go-past times (the time from first fixating a word to moving past it) on word *n* were longer for Ungrammatical than Same-Context [gaze: 38ms difference, t=2.7; go-past: 44ms, t=2.2] and Same-*n*+2 [gaze: 27ms, t=1.8 (marg.); go-past: 43ms, t=2.1]. In addition, readers regressed out of word *n* more for Ungrammatical than Same-Context [z=1.8 (marg.)] and Same-*n*+2 [z=2.2].

Conclusion. Our finding of more processing difficulty on word *n* in the Ungrammatical condition than in either Grammatical suggests that at this point comprehenders can already evaluate the grammatical fit of word *n*+2 into its context. Two possible interpretations are: (i) that comprehenders identify and then attempt to integrate word *n*+2 into the current parse; and (ii) after integrating word *n*+1 into the current parse, comprehenders have strong expectations about the grammatical class of word *n*+2 (e.g., determiner), which lead to strong expectations about its length (e.g., short), which are violated by the long word to the right. On view (ii), these results are highly consistent with reports of neural correlates of word grammaticality violations at 100ms post-onset (Dikker et al., 2009). On either view, these results both demonstrate comprehenders' extreme speed and eagerness in constructing grammatical representations, and fit well the growing view of comprehension as a highly predictive process.

Example: (target word *n* in *italics*; onset of ungrammaticality in **bold**)

Ungrammatical:	The youth in the *class* and **skimmed** …
Grammatical, Same-Context:	The youth in the *class* and the …
Grammatical, Same-*n*+2:	The youth detested the *class* and skimmed …

The relationship between regressive saccades and the P600 effect: Evidence from concurrent eye movement and EEG recordings

Paul Metzner (University of Potsdam), Titus von der Malsburg (St John's College, University of Oxford), Shravan Vasishth (University of Potsdam), Frank Rösler (University of Hamburg)
paul.metzner@uni-potsdam.de

Reading studies using eyetracking have found that word integration difficulties are reflected in elevated reading times and regressive eye movements (Clifton, Staub, & Rayner, 2007). Studies measuring event-related brain potentials (ERPs) have found that semantic and syntactic integration difficulties also elicit characteristic brain responses: N400 effects for semantic (Kutas & Hillyard, 1980) and P600 effects for syntactic integration (Osterhout & Holcomb, 1992). For two reasons, it is not clear how eyetracking and ERP effects map onto each other. First, no regressions occur in ERP designs where sentences are presented word-by-word (RSVP). Second, eye movements in natural reading produce artifacts in the EEG, which until recently complicated the analysis considerably. Thus, the precise relation of eye movements and ERPs is unclear.

To address this issue, we conducted a study of both types of integration difficulty in which we recorded eye movements and brain potentials concurrently and used ICA (Jung et al., 2000) to correct the eye movement artifacts in the EEG. The materials in our study followed a design developed by Hagoort (2003):

(1) Der/**Die** verfallene/*neugierige* Bauernhof braucht eine Renovierung. Er ist…
The$_{[masc]}$/**The**$_{[fem]}$ run-down/*inquisitive* farm needs a renovation. It is…

(2) Der erfahrene Star spielt die/**das** schwierige/*elektrische* Rolle. Er überzeugt…
The experienced star plays the$_{[fem]}$/**the**$_{[neut]}$ difficult/*electric* role. He convinces…

Sentence-internally (1) or sentence-finally (2), sentences could contain a gender mismatch between determiner and noun (syntactic violation, bold face for violating determiner, critical noun underlined) or an adjective that did not match with the noun semantically (semantic violation, italics). To have a baseline for the natural reading ERPs, we presented the same sentences using RSVP and randomly assigned participants (N=72) to one of the two presentation modes.

We replicated the results of the Hagoort study with RSVP (N=24): Syntactic violations elicited a P600 in sentence-internal position and an N400-P600 response in sentence-final position, semantic violations an N400-P600 in sentence-internal and an N400 in sentence-final position. The ERPs obtained from natural reading (N=48) were qualitatively similar when we analyzed the entire data set. When we analyzed only violation trials with regression from the critical noun, we found a P600 in sentence-internal position and an N400-P600 in sentence-final position for syntactic and semantic violations. In contrast, we found no P600 in either position and for either syntactic or semantic violations when no regression occurred. In sentence-final position, we instead found a sustained centro-parietal negativity for both violations.

This demonstrates two things: First, the ERP results from natural reading replicate the results from RSVP, which is further evidence for the method's validity (see also Dimigen et al., 2011; Kretzschmar et al., 2009). Second, the results show that the P600 effect is associated with regressions in natural reading (see also Dimigen et al., 2007), lending support to the idea that recovery processes are accompanied by a P600 effect. The fact that such an effect was absent in trials without regression suggests that the parsing system pursued a qualitatively different processing strategy that may not involve attempted recovery. Thus, the benefit of the coregistration approach is two-fold: First, it provides us with a behavioral signal that enables us to better understand ERPs. Second, the ERP results allow us to attach functional interpretations to eye movement phenomena like regressions.

Unforced revision in the processing of relative clause association ambiguity in Japanese

Toshiyuki Yamada, Manabu Arai, & Yuki Hirose (The University of Tokyo)
t_yamada@phiz.c.u-tokyo.ac.jp

Previous studies on relative clause (RC) association ambiguity in Japanese (RC N1-Gen N2) have shown mixed results: initial local association (LA) and eventual non-LA preferences (Kamide & Mitchell 1997; Miyamoto et al. 2004). One possible cause is the difference between on-line and off-line measures. This suggests a possibility that Japanese comprehenders adopt LA as an initial analysis but revise it to non-LA when N2 is encountered. To explore this, the current study examined the real-time processing of RC ambiguity using eye-tracking technique.

Experiment 1 investigated whether people revise the initial LA analysis to non-LA at N2. (1a-d) are globally ambiguous up to N2 (*sobo* 'grandmother') but (1a-b) force the LA analysis due to semantic incompatibility of non-LA at the main predicate (i.e., adverb and/or verb). We also manipulated thematic fit between RC and N1: N1 is more plausible in (1a,c) compared to (1b,d) (McRae et al. 1998). If non-LA is indeed adopted at N2, we predict the following: (i) Reading times at N2 would be longer in (1a,c) than in (1b,d) because the revision of the strongly committed LA analysis in (1a,c) would be more costly. (ii) Greater difficulty should occur at the main predicate in LA-forced (1a-b) compared to (1c-d). Twenty-eight participants' eye-movements in reading of twenty-four target items were recorded. First pass times showed that the main adverb in (1a-b) took longer to read than (1c-d) ($p < .05$). This most likely reflects ambiguity advantage (Traxler et al. 1998), confirming (ii). Furthermore, there was an interaction between association bias and thematic fit in go-past times at the main adverb. Further analysis showed a significant difference in (1a-b) (1095 ms, 883 ms) ($p < .05$) but no difference in (1c-d) (911 ms, 1027 ms). This is likely to reflect spill-over from the preceding N2 region, supporting (i).

Experiment 2 further examined the strategy of revising the initial LA analysis to non-LA at N2. Using inanimate N2 (*mannenhitu* 'fountain pen'), we created LA-forced (2a-b) (N1 is the only possible head for RC) and compared them with globally ambiguous (2c-d). Thematic fit was again manipulated. If non-LA is indeed adopted at N2, we predict the following: (iii) The reading time in early measures at N2 would be longer in (2b) than in (2a) because the revision of the strongly committed LA analysis in (2a) would be cancelled for highly implausible N2. (iv) Reading times in late measures at N2 would be longer in (2c-d) compared to (2a-b) because both N1 and N2 are potential heads for RC in (2c-d). We tested twenty-eight new participants using the same procedure. The interaction between association bias and thematic fit was observed in first pass times at N2. Further analysis showed a significant difference in (2a-b) (338 ms, 381 ms) ($p < .05$) but no difference in (2c-d) (358 ms, 346 ms), which supports (iii). Furthermore, second pass times at N2 were longer in (2c-d) than in (2a-b) ($p < .05$), which confirms (iv).

Our study demonstrated that Japanese comprehenders make an unforced (i.e., not syntactically required) revision from the initial LA analysis to non-LA on encountering N2, contrary to Revision as Last Resort (Fodor & Frazier 1980), and that this revision can be cancelled when the N1 is thematically highly plausible as a head for RC and the N2 is rather implausible. The results together motivate a parser that is willing to weigh an alternative analysis as it becomes available even after the initial analysis is adopted and to revise it for a maximally coherent interpretation.

(1) a/b. [$_{RC}$ Sakka-ga syosai-de mendansiteiru] hensyūsya/dansei-no sobo-ga hokkyoku-ni ryokōsiteiru.
writer-Nom study room-in talking editor/man-Gen grandmother-Nom North Pole-to traveling
'The grandmother of the editor/man that the writer is talking with in the study room is traveling to the North Pole.'

c/d. [$_{RC}$ Sakka-ga syosai-de mendansiteiru] hensyūsya/dansei-no sobo-ga āmuchea-ni kosikaketeiru.
writer-Nom study room-in talking editor/man-Gen grandmother-Nom armchair-on sitting
'The grandmother of the editor/man that the writer is talking with in the study room is sitting on the armchair.'

(2) a/b. [$_{RC}$ Sakka-ga syosai-de mendansiteiru] hensyūsya/dansei-no mannenhitu-ga tēburu-kara otita.
writer-Nom study room-in talking editor/man-Gen fountain pen-Nom table-from fell
'The fountain pen of the editor/man that the writer was talking with in the study room fell from the table.'

c/d. [$_{RC}$ Sakka-ga syosai-de mitumeteiru] hensyūsya/dansei-no mannenhitu-ga tēburu-kara otita.
writer-Nom study room-in staring editor/man-Gen fountain pen-Nom table-from fell
'The fountain pen of the editor/man that the writer was staring at in the study room fell from the table.'

Ambiguity Affects Mandarin Relative Clause Processing

Yaling Hsiao (UW-Madison), Jinman Li (SHUFE) & Maryellen MacDonald (UW-Madison)
yhsiao5@wisc.edu

Unlike patterns in many other languages, Mandarin object relative clauses (ORs) may be easier than subject relatives (SRs), making them important in theories of how syntactic complexity affects relative clause (RC) interpretation. Such effects are complex, with reports of replications, failures to replicate, and evidence that the effect holds for RCs modifying main clause subjects but not objects. Meanwhile, studies in English suggest that RC difficulty stems from ambiguities in RCs rather than SR/OR syntactic complexity. Corpus analyses show that Mandarin RCs contain significant ambiguities as a function of RC word order and main clause modifying position (e.g. the "N V" word order in ORs induces different degrees of ambiguity when modifying different main clause positions), but to date no work has tested whether difficulty in Mandarin RC ambiguity resolution could supplant or supplement accounts resting on syntactic complexity. This study aims to investigate ambiguity in Mandarin RCs.

Exp.1: A self-paced reading study (N=32) crossed RC type and RC modifying position (Table1; not shown are preceding context & post-RC continuations, avoiding sentence wrap-up effects that plague some previous studies). Length-adjusted log RTs were fitted to mixed-effect models. As shown in Fig1, subject-mod RCs were read faster than object-mod RCs throughout. ORs were easier than SRs at W2 (replicating Hsiao & Gibson, 2003; Gibson & Wu, 2013), whereas SRs were easier than ORs at the head noun (replicating Lin, 2006; Vasishth et al., 2013). Exp. 2 tests if these complex patterns owe to varying points of ambiguity in RCs.

Table1	**Subject Relative (SR)** (W1=praise, W2=director)	**Object Relative (OR)** (W1=director, W2=praise)
Subject-mod	…[称赞 导演 的] 演员 感谢了 编剧… …[praise director DE] actor thanked screenwriter… the actor that praised the director thanked the screenwriter	…[导演 称赞 的] 演员 感谢了 编剧… …[director praise DE] actor thanked screenwriter… the actor that the director praised thanked the screenwriter
Object-mod	…编剧 感谢了 [称赞 导演 的] 演员… …screenwriter thanked [praised director DE] actor… the screenwriter thanked the actor that praised the director	…编剧 感谢了 [导演 称赞 的] 演员… …screenwriter thanked [director praised DE] actor… the screenwriter thanked the actor that the director praised

Exp.2: Gennari & MacDonald (2008) used "gated" sentence completions to track English readers' expectations and found that ambiguity (i.e., the rate of non-RC completions for an RC-beginning sentence fragment) predicted reading times at points that forced the RC interpretation. We collected completions for Exp.1 items at two points (N=56). Gate1 was predicted to be an early locus of high ambiguity with high possibility of garden path (fragments up to W2 for subject-mod RCs, ending at W1 for object-mod). Gate2 for both modifying positions ended at the relativizer DE, which strongly promotes an RC. Fig.2 confirmed these predictions. We added 2 new factors to Exp.1's multilevel model at the head noun: Gate2 RC completion rate, and a Gate2-Gate1 difference score (reflecting the degree to which early alternative interpretations are contradicted at DE, i.e. the slope of the lines in Fig.2). Whereas RC type was robust in the Exp.1 analyses, it no longer predicted RTs with the addition of the completion data. This shows for the first time that RC difficulty in Mandarin stems from ambiguity resolution difficulty, and syntactic structure by itself does not add to an account of comprehension difficulty. We discuss ambiguity effects at other word positions and implications for theories of RC interpretation.

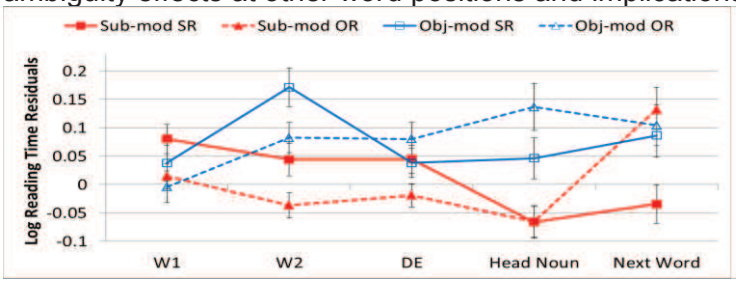


Fig.1 Length-adjusted log reading time residuals in Exp.1 (W1: RC verb for SRs & RC noun for ORs; W2: noun for SRs & verb for ORs, DE=relativizer).

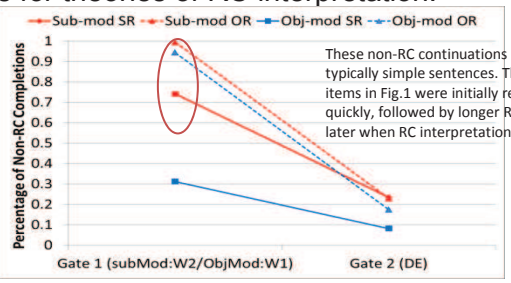


Fig. 2 The percentage of non-RC completions at the two Gates in Exp.2

Underspecification in RC attachment: A speed-accuracy tradeoff analysis

Pavel Logacev & Shravan Vasishth (University of Potsdam)
pavel.logacev@gmail.com

Background Traxler et al. (1998) found that sentences with ambiguous RC attachment are read faster than their unambiguous counterparts. One explanation for this so-called *ambiguity advantage* is the good-enough account by Swets et al. (2008): They argue that readers underspecify the representation of ambiguous sentences unless disambiguation is required by task demands. Thus, ambiguous sentences are processed faster because the RC does not need to be attached. When disambiguation is provided, however, readers assign sentences full structure. This additional operation results in slowed reading of unambiguous sentences. Assuming that a minimal amount of time is required to complete RC attachment, the underspecification account predicts that ambiguous sentences should not only be processed faster than unambiguous sentences on average, the minimum time to process an ambiguous RC should be shorter as well. We tested this prediction in an experiment using the speed-accuracy-tradeoff (SAT) paradigm (e.g., McElree, 1993).

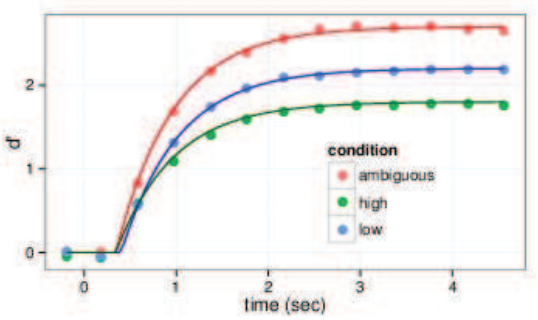
Method Twenty participants judged the acceptability of sentences such as (1) at different lags after the presentation of the last phrase (the RC). Attachment was manipulated by means of gender agreement with the relative pronoun.

(1) a. Was dachte **der Manager** **der Sängerin,** [$_{RC}$ **der/die** schwieg?] (high/low)
*What thought the manager.**MASC** of the singer.**FEM**, who.**MASC/FEM** was silent*

b. Was dachte **der Manager** **des Sängers,** [$_{RC}$ **der** schwieg?] (ambiguous)
*What thought the manager.**MASC** of the singer.**MASC**, who.**MASC** was silent*
'What did the manager of the singer who was silent think?'

Predictions The difference in minimum completion times predicted by the underspecification account means that the time at which participants' accuracy departed from chance should be shorter for ambiguous than unambiguous sentences. In other words: The intercept of the SAT function should be lower for ambiguous than for unambiguous sentences.

Results We estimated the parameters for a range of models ranging from minimal (1 asymptote, 1 rate, 1 intercept) to fully saturated (3 of each) using MLE. The figure shows the average sensitivity (d') overlaid with the best fit of the fully saturated model. The estimated asymptotes for the ambiguous, high and low conditions were 2.7, 2.2, and 1.8, while the intercept estimates were: 343 ms, 400 ms, 343 ms. BIC model-selection based on by-participant model fits preferred 3 asymptote-1 intercept models.The lack of differences in intercept is not compatible with the underspecification account. Moreover, the difference in asymptotes suggests that the ambiguity advantage may correspond to more parser failures in unambiguous conditions.



References

McElree (1993). The locus of lexical preference effects in sentence comprehension. JML, 32.
Swets, Desmet, Clifton, & Ferreira (2008). Underspecification of syntactic ambiguities.. Memory & Cognition, 36.
Traxler, Pickering, Clifton, & Clifton (1998). Adjunct attachment is not a form of lexical ambiguity resolution. JML, 39.

30 **Linear order effects in agreement: Evidence from English *wh*-questions**

Brian Dillon, Joshua Levy, Adrian Staub & Charles Clifton, Jr (UMass Amherst)
brian@linguist.umass.edu

Production studies have suggested that agreement attraction effects are due primarily to structural intervention of a non-subject NP, as opposed to the linear word order of this element with respect to the subject and verb [1,2]. However, it has been argued that linear order effects on agreement do arise in certain circumstances [3]. In the present study we demonstrate a novel effect of linear order, arising in sentences such as (1).

(1) **Which flowers is the gardener planting?**

We find that the mismatch in number between the fronted wh-object and the auxiliary has clear effects on both off-line acceptability judgments (Exp 1,2) and on-line processing (Exp 3). The experiments suggest, furthermore, that these effects are not attributable to an initial misanalysis of the fronted wh-object as subject (c.f. [4]): they arise even when this misanalysis is not possible (Exp 2) and persist in online measures past the point of disambiguation (Exp 3). We suggest that the results implicate a heuristic process for computing agreement in English: comprehenders appear to treat a noun in the constituent to the left of an inflected verb as a potential agreement controller, regardless of whether it is an actual subject position.

Expt 1 (n = 40) employed an offline 1-7 acceptability task with a 2x2x2 design, crossing number of the fronted *wh*-object (sing/pl), number of the subject (sing/pl), and number of the auxiliary (sing/pl). LME modeling of the results revealed a three way interaction (t = 6.4, p < .0001): number mismatch between the fronted object and the auxiliary reduced acceptability of grammatical sentences, and this effect was stronger when the object was plural. In addition, number match between these elements increased acceptability of ungrammatical sentences. Expt 2 (n =40) used pied piping to block the initial subject misanalysis, crossing number of the subject (sing/pl), number of the auxiliary (sing/pl), and location of the preposition (in-situ/fronted). The results reveal an interaction between fronting and auxiliary number (t = 3.3, p < .001), such that the penalty for the object/auxiliary mismatch is lessened with a fronted preposition. However, planned comparisons reveal a significant drop in acceptability due to aux/object mismatch even for pied-piped conditions (**2b** v **2f**: 4.6 vs 5.6), ruling out accounts that attribute this effect entirely to misanalysis of the fronted object. Two additional acceptability studies confirm that the effects of object-aux match are present with pronominal subjects, and are also present when the number-inflected object is not linearly adjacent to the auxiliary.

An in-progress eye-tracking experiment (Expt 3; n =35) investigates the online processing of grammatical sentences **1a,b** and **1e,f**, crossing object-auxiliary MATCH with AUX number. LME modeling reveals a significant effect of MATCH beginning on a region including the auxiliary and the subsequent determiner in regressions out (Wald's z = -2.3, p < 0.05) and go-past times (t = -3.3, p < .001). When the auxiliary is singular, this effect persists as late as the post-subject spillover region, while it is resolved earlier when the auxiliary is plural (MATCH×AUX interaction at spillover: regressions out, Wald's z = -3.1, p < 0.005; go-past t = -1.96, p < 0.05).

Expt 1, mean ratings	**Expt 2, mean ratings**
a. Which flowers is the gardener..: 4.2	**a. Which players is the coach talking to? 4.2**
b. Which flower is the gardener..: 6.7	**b. To which players is the coach talking? 4.6**
c. *Which flowers is the gardeners..: 3.2	**c. *Which players is the coaches talking to? 2.7**
d. *Which flower is the gardeners..: 3.6	**d. *To which players is the coaches talking? 3.2**
e. Which flower are the gardeners..: 5.6	**e. Which players are the coaches talking to? 6.2**
f. Which flowers are the gardeners..: 6.6	**f. To which players are the coaches talking? 5.6**
g. *Which flower are the gardener..: 2.6	**g. *Which players are the coach talking to? 3.5**
h. *Which flowers are the gardener..: 4.5	**h. *To which players are the coach talking? 3.5**

Expt 3, [*is/are the*] region	**Sing: go-past**	**p reg**	**Pl: go-past**	**p reg**
Match	**356 (+/- 32)**	**0.21**	**322 (+/- 22)**	**0.13**
Mismatch	**465 (+/- 40)**	**0.27**	**488 (+/-50)**	**0.26**

[1] Eberhard, Cutting & Bock. (2005). *Psychological Review* [2] Vigliocco & Nicol. (1998). *Cognition*. [3] Haskell & McDonald, 2005. *JEP: LMC* [4] Gorrell. 1998. In Fodor & Ferreira, (eds), *Reanalysis in Sentence Processing*.

The influence of *verb-specific featural restrictions*, *word associations*, and *production-based mechanisms* on language-mediated anticipatory eye movements

Florian Hintz, Antje S. Meyer, & Falk Huettig (*Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*
florian.hintz@mpi.nl

Why is language comprehension so fast? One answer to this question might be: Because humans can predict upcoming language input. While there is a growing body of literature on experimental studies exploring the contents of the comprehenders' predictions, empirical evidence for proposed mechanisms underlying anticipation in language processing is sparse. One proposal is that listeners or readers imagine what they would say next as a speaker, using their language production system cf. [1]. Those theories essentially suggest that predicting a word is like producing a word. Alternatively, some theoretical frameworks propose that prediction happens on the basis of simple word associations [2]. That is, on processing a given word, activation of that mental representation automatically spreads to associated concepts. Lastly, priming experiments have shown that verbs, for example, activate typical patient nouns but only if those meet verb-specific featural restrictions [3]. Crucially, the latter two accounts predict that differences in association strength between concepts or differences in the fit of featural restrictions modulate the degree of anticipation.

This study investigated the relative contribution of those three mechanisms to language-mediated anticipatory eye movements. In a visual world experiment, participants (Exp 1 = 61, Exp 2 = 60, Exp 3 = 60) looked at sets of four objects and listened to predictive sentences (N = 40) such as "The man peels an apple*"* or non-predictive sentences such as "The man draws an apple". The sentences varied in verb-noun association strength (mean = .37, range .09 to .77) as assessed in a free association task. Moreover, we quantified verb-specific featural restrictions using a typicality rating (e.g., "How common is it for an apple to be peeled?"). On predictive trials, only the target object fitted the semantic constraints of the verb. When the objects were presented before the verb was heard (Experiments 1 and 2), participants, as expected, looked at the target object before it was mentioned. In both experiments, the likelihood of predictive looks correlated positively with participants' production fluency (Exp 1: $r = .332$, $p = .011$; Exp 2: $r = .187$, $p = .16$) as measured in a separate verbal fluency task (categories: animals, professions; letters: L, M). When participants were given only a short preview of the display (after the verb was heard, Experiment 3), the correlation between the likelihood of predictive looks and participants' production fluency was absent ($r = -.061$, $p = .645$). In all three experiments, anticipatory fixations of the target objects correlated positively with the items' typicality rating (residualized for association strength; Exp 1: $r = .451$, $p = .004$; Exp 2: $r = .352$, $p = .028$; Exp 3: $r = .366$; $p = .022$). Verb-noun association strength did not modulate anticipatory eye movements in any of the experiments.

We conclude that verb-specific featural restrictions are a robust contributor to language-mediated anticipatory eye movements. The involvement of production-based mechanisms seems to be constrained to situations where visual input is present when predictions are made. We conjecture that the presence of visual objects stimulates the engagement of the production system as comprehenders might be more inclined to imagine how they would finish the sentence given the visual objects. Word associations, did not influence anticipation in our experiments, but may do so in other settings. We suggest that the different mechanisms (e.g., featural restrictions and production-based mechanisms) play different roles depending on the situational context. Theories of prediction have to take *situational context* (e.g., the amount of visual input) into account.

References: [1] Pickering, M.J. & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36, 329-347. [2] Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, 119, 23–42. [3] Ferretti, T. R., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44, 516–547.

Spillover frequency effects in a sequential sampling model of reading

Michael Shvartsman, Richard L. Lewis & Satinder Singh (University of Michigan)
mshvarts@umich.edu

Overview. Words that follow high-frequency words are read more quickly [1]. This robust spillover effect is of interest both for understanding eye movement control, and for relating eye movement measures to underlying language processes. The multihypothesis sequential probability ratio test (MSPRT, [2]) is a statistical decision procedure for selecting the most probable hypothesis from a set (e.g. words) given a sequence of incrementally obtained evidence. The test stops obtaining evidence and selects a hypothesis (i.e. identifies a word) when the highest posterior probability crosses a threshold. The MSPRT is attractive as a basis for saccadic control in reading: it provides a rational explanation of the logarithmic frequency effect without building it in [3], it recovers lexical neighborhood effects [3], and it is consistent with neural evidence from simple saccadic decision tasks [4]. A potential problem is that the threshold may put low and high frequency words on equal footing for future processes, making it impossible to obtain spillover effects of frequency. We show that this is not the case, and provide a novel explanation for spillover effects in a sequential sampling framework.

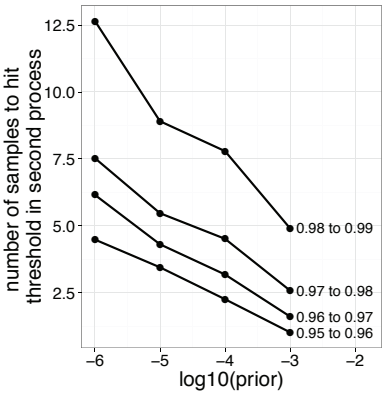


Figure 1. Steps-to-threshold in an MSPRT that takes as prior the posterior of a previous MSPRT. The horizontal axis is the prior of the target at the start of the first MSPRT.

MSPRT posterior at stop is sensitive to prior. In the MSPRT, the mean stopping posterior probability for a hypothesis is sensitive to its initial prior: lower priors lead to lower mean stopping posteriors. This is because incorrect decisions are likelier to happen at lower priors, and drive the stopping probability down on average for those priors. Fig. 1 shows the time-to-threshold of a *second* sampling process as a function of the prior probability of the target hypothesis (word) for a *first* sampling process. The second MSPRT is sensitive to the original prior despite the "leveling" effects of the first threshold. To the extent that the second MSPRT delays processing on the following word, a spillover frequency effect will arise.

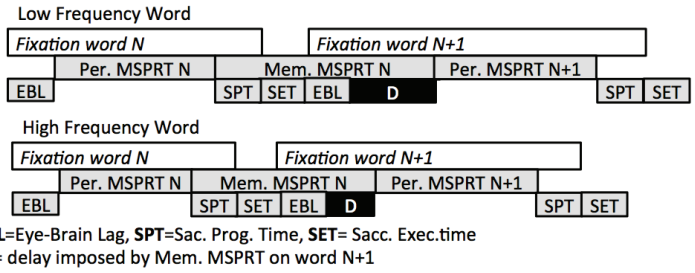


Figure 2. Sequence diagram of overshoot-based spillover model.

A computational model of saccadic control that yields spillover as a function of MSPRT overshoot. We instantiate this explanation in a model (Fig. 2) that uses a serial architecture inspired by E-Z Reader [5]: information arrives from the eye with some delay (eye-brain lag), and enters an MSPRT sampler. When this sampler reaches threshold, saccade planning is initiated, followed by saccade execution and the eye-brain lag for the next word. Meantime, processing of the previous word continues via a second, memory-driven MSPRT that rises to a second threshold; this makes efficient use of oculomotor slack time but potentially delays processing on the next word. The model recovers spillover frequency effects of a magnitude consistent with the literature (SFD on words following LF words: 266ms; on HF words: 281ms; HF and LF determined by median split). This is the first demonstration of spillover in a sequential sampling framework, and provides a link for how a range of higher-level processes may be coordinated with saccadic control.

References [1] Inhoff & Rayner (1986). [2] Baum and Veeravalli (1994). [3] Norris (2006) [4] Bogacz (2007). [5] Reichle, Pollatsek & Rayner (2012).

Auditory Perceptual Simulation during Silent Reading

Peiyun Zhou (University of Illinois at Urbana-Champaign) & Kiel Christianson (University of Illinois at Urbana-Champaign, Beckman Institute for Advanced Science and Technology)
pzhou5@illinois.edu

Embodied cognition theory proposes that semantic meaning derives from the activation of perceptual features associated with physical experiences and sensations (Barsalou, 1999; Barsalou & Hale, 1993; Glenberg, & Robertson, 2000; Zwaan, 2008). Studies have demonstrated that language processing elicits perceptual simulation: readers perceptually simulate the motion, action, odor, and other information described in text during reading (Bergen, &Wheeler, 2010; Glenberg, & Kaschak, 2002; Hubbard, 2010; Louwerse, & Jeuniaux, 2010; Pecher et al., 2003, 2004). Thus, during language processing, readers “see” depicted objects and events, “smell” described odors, “hear” depicted voices, and “act” described motions in their minds. We report three eye-tracking experiments exploring the effects of *auditory* perceptual simulation (APS) on sentence processing and comprehension during silent reading.

Method: In Experiment 1, to induce 80 participants to perceptually simulate the speech of one native and one nonnative English speaker, we provided them with examples of each speaker's voice. Four 500-word passages were recorded as they were read aloud by a native speaker of American English and a Chinese nonnative English speaker. These recordings, which differed only in speaker accent and rate (the non-native speaker spoke more slowly), were played for participants as they were shown a photograph of each speaker. The two speakers’ photos were subsequently used (counterbalanced) as the drift-correct picture prior to each sentence presented in the experiment to cue the "speaker's" voice as participants silently read the sentences, which were all in quotations marks. 48 target sentences and 96 fillers were read. The target sentences included four conditions: (1a) plausible subject-relative (SRC-plausible); (1b) implausible subject-relative (SRC-implausible); (1c) plausible object-relative (ORC-plausible); (1d) implausible object-relative (ORC-implausible). This resulted in a 2(structure) x 2(plausibility) x 2("speaker") fully factorial design. After each sentence, participants answered a paraphrase verification question (1e), the correct answer to which was always "yes". Half way through the experiment, two more recordings of the native and nonnative speakers' voices were played to remind participants what they sounded like. Participants’ eye movements were monitored, and reading times and responses were recorded using an EyeLink 1000 eye tracker. Also, a social attractiveness scale was administered to measure participants' attitudes toward the "speakers'" speech.

(1) a. “The bird that ate the worm was small.” (SRC-plausible)
b. “The worm that ate the bird was small.” (SRC-implausible)
c. “The worm that the bird ate was small.” (ORC-plausible)
d. “The bird that the worm ate was small.” (ORC-implausible)
e. The worm/bird ate the bird/worm. The worm/bird was small. (Yes No)

In Experiment 2, participants read the same sentences as Experiment 1. Two recordings of speakers’ names, instead of two speakers’ photos, were used before each sentence as speech cues. In Experiment 3, 40 participants were recruited to read the same sentences normally, without quotation marks and without any APS cue, in the same procedure.

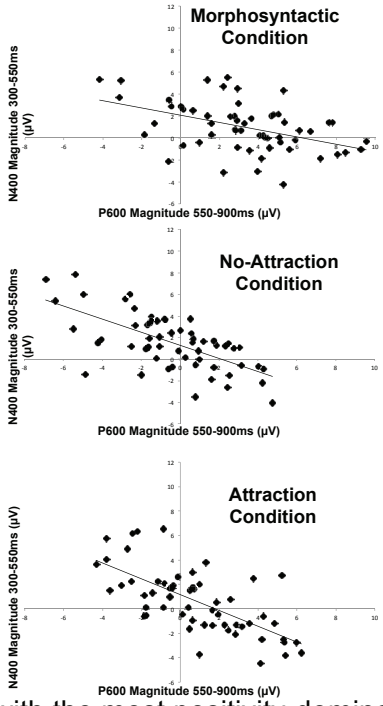
Results: In E1 and E2, participants read sentences attributed to the native speaker faster than those attributed to the nonnative speaker, consistent with the speakers' respective speech rates. In E1 but not E2 participants were more likely to miscomprehend sentences after perceptually simulating nonnative speech. Importantly, APS of either native or nonnative speech in E1 and E2 yielded higher accuracy than no APS in E3, including a 20% improvement in the ORC-implausible condition. Reading speed in E1 was statistically faster in both native and nonnative conditions than in E3, and speed in E2 was also numerically faster than in E3. Taken together, the results indicate that APS facilitated sentence processing and improved comprehension. Broader implications of APS in educational settings will also be discussed.

Gradience and individual differences in processing syntactic, semantic and thematic information: ERP evidence

Darren Tanner (University of Illinois) & Janet G. van Hell (Pennsylvania State University)
dstanner@gmail.com

Recent studies of sentence comprehension using event-related potentials (ERPs) have provided conflicting evidence regarding how combinatorial constraints, such as verb-argument animacy restrictions, interact with lexical-semantic associations between words. Kim and Osterhout (2005) suggest that lexical-semantic associations between words are processed independently of morphosyntactic cues to thematic role assignment. They demonstrate that progressive verbs following inanimate subject noun phrases elicit P600 effects when the subject NP makes a plausible Theme of the verb, but an N400 when the subject was an implausible Theme. On the other hand, Kuperberg and colleagues (2007; Paczynski & Kuperberg, 2012; see also Stroud & Phillips, 2012) show that all animacy violations elicit P600 effects, which are not modulated by semantic associations. On this account, animacy restrictions are a hard combinatorial constraint, where the P600 effect reflects categorical processing of a propositional impossibility. However, important caveats to all of this work are the assumptions that individuals respond categorically with N400 and/or P600 effects, that all individuals process these violations similarly, and that the grand mean reflects this normative brain response across all participants. We show here that each of these assumptions needs reconsideration.

We investigated the processing of morphosyntax and the interaction between animacy restrictions and semantic associations through the lens of individual differences. We show that not only are brain responses to animacy violations modulated by semantic attraction, but that these effects are gradient across individuals. Fifty eight monolingual English speakers read sentences that were well-formed (*The broken television was repaired…*), morphosyntactically anomalous (*The broken television was repairs…*), contained an animacy violation with no semantic attraction between subject and verb (*The hearty meal was repairing…*), or an animacy violation with a semantic attraction between subject and verb (*The broken television was repairing…*). Grand mean results showed a reliable biphasic N400/P600 response in the morphosyntactic violation condition, a reliable N400 in the No-attraction condition, and a small, marginally reliable biphasic N400/P600 response in the Attraction condition. However, inspection of individuals' ERP responses showed that in no condition did the grand mean accurately reflect the distribution of most participants' brain responses. Instead, participants' brain responses varied along a continuum between negativity-only (upper left quadrants in the figures), biphasic (upper right quadrants), and positivity-only responses (lower right quadrants). Moreover, the midpoint of the distribution varied as a function of the number and type of cues signaling the violation, with the most positivity-dominance in the Morphosyntactic violation condition, most negativity-dominance in the No-Attraction condition, and equal response dominance in the Attraction condition, where both semantic and morphological cues signaled the violation. These results show that brain response modulations by syntactic, semantic and thematic cues are not categorical, but gradient. They additionally show that there are marked, but systematic individual difference in how people process this information. Individuals fall along a continuum between N400 and P600-dominant responses, though location along this continuum varies as a function of the number and type of cues signaling the violation.



The Contribution of Form and Meaning to the Processing of Careful and Casual Speech.

Kevin McGowan, Meghan Sumner, Annette D'Onofrio, and Teresa Pratt (Stanford University)
sumner@stanford.edu

Variation in speech abounds. Listeners experience a casual speech style much more often than a careful style (similar to citation form). Various studies pit words produced in these two styles against each other in search of evidence that one forms the basis of linguistic representations –investigating whether linguistic representations match frequently observed utterances (as in casual speech) or utterances that are infrequent but ideal (as in careful speech). Many of these studies focus on word-level effects and compare a natural token to either a manipulated token or an illicit production of a word. However, when different naturally-uttered, licit productions of words are compared, listeners recognize both frequent (casual) and infrequent (careful) productions equally well. This equivalence is problematic for theories that depend heavily on observed frequency as a predictor of listener ability. To address this equivalence, this study shifts from words to sentences, testing a hypothesis recently proposed by Sumner (2013): As the typical production pattern, casual speech processing depends more on top-down, semantic predictability and, as the atypical but idealized production pattern, careful speech processing depends more on bottom-up, form-based mapping of a speech signal to linguistic representations. This should result in dense and weakly encoded episodic clusters for typical productions but sparse and strongly encoded episodic clusters for atypical forms that are standard –predicting equivalence by two very different means.

We present data from two cross-modal priming studies in which listeners are presented with casually- or carefully-articulated sentences followed by a visual target. We use SPIN test sentences to manipulate semantic predictability (Bilger 1984). In Exp. 1, we investigate responses dependent on speech style (casual vs. careful), semantic predictability (predictable: The cow was milked in the **barn**; unpredictable: I am thinking about the **barn**), and relatedness of the probe (related: **barn**, unrelated: **coin**). Speech style was a between-subject factor and visual targets were presented 100 msec after the offset of the auditory sentence. This conflated form repetition and semantic predictability and served as a first step to investigate both the main manipulation of relatedness and the effect of semantic predictability dependent on spoken speech style. Overall, we found that related targets were identified faster than unrelated targets (β=5.95; t=137. 13; p<0.001), and this effect is stronger for careful than casual (β=-0.03; t=-2.67; p<0.01). Semantically predictable targets were recognized more quickly than unpredictable targets in the casual condition, but not in the careful condition (β=-0.038; t=-2.36; p<0.05). These findings suggest that the different speech conditions may have prompted more form-driven responses to careful speech and more meaning-based responses to casual.

To decouple form from meaning, we moved the target probe earlier in the sentence in Exp. 2. Probe placement explicitly investigated the dimension of time in speech processing: While careful speech has an expanded vowel space and more idealized sound productions generally, it also takes longer to utter than casual speech. In Exp. 2 we ask whether, for semantically-predictable sentences, related targets are recognized more quickly than unrelated targets when the target probe occurs at the end of the casually-articulated word or at roughly half of a carefully articulated word (the exact point was determined by the duration of each casually-articulated token). We again find a main effect of relatedness, but this is driven by the careful speech condition alone, with no facilitation for the casual speech condition (β=0.04; t=3.31; p<0.01). Critically, though, targets following casually-articulated speech were recognized much more quickly than those following carefully-articulated speech (β=-0.11; t=-2.08; p<0.05). RTs were nearly 100 msec faster for targets following casual than careful speech, suggestive of a floor effect in this much more difficult task. Preliminary analysis of a visual world eye-tracknig paradigm designed to replicate these studies using a more sensitive task appears to support this interpretation. Taken together, our data suggest that equivalence across speech styles results from differential dependence on form- and meaning-based information.

Expectation-adaptation in the incremental interpretation of contrastive prosody

Chigusa Kurumada, Meredith Brown, Sarah Bibyk, Daniel Pontillo, & Michael K. Tanenhaus
(University of Rochester)
ckurumada@bcs.rochester.edu

The realization of prosody varies across speakers, accents, and speech conditions (e.g., Ladd, 2008). Listeners must navigate this variability to converge on consistent prosodic interpretations. We investigate whether listeners adapt to speaker-specific realization of prosody based on recent exposure and, if so, whether such adaptation is rapidly integrated with online pragmatic processing. To this end, we investigate contrastive focus, which can signal that pragmatic inference is required to determine speaker meaning (e.g., Ito & Speer, 2008; Pierrehumbert & Hirschberg, 1990; Watson et al., 2008). We find that 1) listeners use contrastive focus to construct and evaluate visually salient contrast sets incrementally; and that 2) changes in the reliability of prosodic cues are reflected in changes in processing time-course.

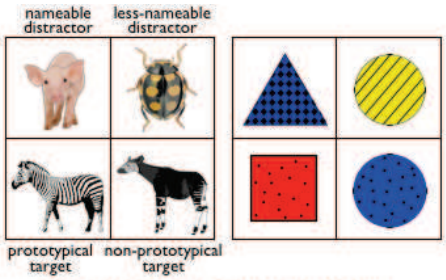


Figure1.Sample visual displays
(Left: main-task, Right: pre-task in Exp.2)

Experiment 1 used the English sentence "It looks like a(n) X" pronounced with either nuclear H* accent on the final noun followed by L-L% (**Noun-focus prosody**) or L+H* on the verb *looks* followed by L-H% (**Verb-focus prosody**). Noun-focus prosody indicates that X is likely to be the identity of the referent, while Verb-focus prosody implies otherwise by evoking contrastive interpretation (e.g., *It LOOKS like a zebra, but it is not one*). We conducted a visual-world experiment with 60 (12 target, 48 filler) trials (Fig. 1, left). Each four-picture display contained a visual contrast pair consisting of a prototypical and a non-prototypical picture of a noun (e.g., zebra) as well as a nameable and a less-nameable distractors. Participants clicked on pictures in response to utterances produced with Noun-focus and Verb-focus prosody. **Results:** A linear mixed regression model examined the effects of prosody (Noun-focus vs. Verb-focus) on log-transformed mean ratios of non-prototypical target fixations (e.g., okapi) to both target fixations (e.g., okapi and zebra). Verb-focus prosody elicited more predictive looks to a non-prototypical target prior to the final noun ($p < .05$), demonstrating rapid generation of a contrastive interpretation based on the L+H* accent.

Experiment 2 tested whether changes in the reliability of contrastive focus affected processing. Experiment 2 was identical to Experiment 1 except for the addition of a pre-task (Fig.1, right) in which two groups of participants heard different acoustic input produced by the same speaker from Experiment 1. In the **high-reliability** condition, participants heard 12 pairs of instructions in which an L+H* on a prenominal modifier was used felicitously (e.g., "Point to the blue circle. Now, point to the YELLOW circle."). In the **low-reliability** condition, 8 out of 12 instances of L+H* were used infelicitously. **Results:** Pre-task exposure affected whether listeners made predictive looks based on L+H* in the main task ($p < .05$). In the high-reliability condition, the L+H* on *looks* resulted in rapid predictive looks to a non-prototypical picture, replicating Experiment 1 ($p < .05$). This early effect was eliminated in the low-reliability pre-task condition ($p > .5$). Additional analysis confirmed that participants in the low-reliability condition still converged on the contrastive interpretation after hearing the L-H% boundary tone, which was not manipulated in the pre-task. This indicates that participants selectively down-weighted L+H* as a cue to a contrastive interpretation rather than discounting prosody entirely.

Listeners thus generate pragmatic expectations by processing prosodic cues incrementally. Furthermore, the reliability of prosodic cues, generalized across different syntactic constructions, can rapidly affect online prosodic processing. We conclude that this rapid adaptation and integration enable listeners to accommodate speaker-specific variability to achieve robust online pragmatic comprehension.

Alternatives in Pragmatic Inference

Judith Degen (Stanford University)
jdegen@stanford.edu

In the face of underspecified utterances, listeners routinely and without much apparent effort make the right kinds of pragmatic inferences about a speaker's intended meaning. I will present a series of studies investigating the processing of one type of inference - scalar implicature - as a way of addressing how listeners perform this remarkable feat.

In particular, I will explore the role of context in the processing of scalar implicatures from "some" to "not all". Contrary to the widely held assumption that scalar implicatures are highly regularized, frequent, and relatively context-independent (e.g., Levinson, 2000), I will argue that they are in fact relatively infrequent and highly context-dependent; both the robustness and the speed with which scalar implicatures from "some" to "not all" are computed are modulated by the probabilistic support that the implicature receives from multiple contextual cues. I will present evidence that scalar implicatures are especially sensitive to a) the naturalness or expectedness of both scalar and non-scalar alternative utterances the speaker could have produced, but didn't; and b) the contextual Question Under Discussion that modulates the relevance of the scalar alternative.

In this context I will present a novel contextualist account of scalar implicature processing that has roots in both constraint-based and information-theoretic accounts of language processing (Degen & Tanenhaus, to appear) and that provides a unified explanation for a) the varying robustness of scalar implicatures across different contexts (Bott & Noveck, 2004; Degen, 2013; van Tiel et al., in prep), b) the varying speed of scalar implicatures across different contexts (Huang & Snedeker, 2009, 2011; Grodner et al., 2010), and c) the speed and efficiency of communication.

References

Bott, L., & Noveck, I. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language*, *51*(3), 437–457.

Degen, J. (2013). *Alternatives in Pragmatic Reasoning*. University of Rochester.

Degen, J., & Tanenhaus, M. K. (to appear). Processing scalar implicature: A Constraint-Based approach. *Cognitive Science*.

Grodner, D. J., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition*, *116*(1), 42–55.

Huang, Y. T., & Snedeker, J. (2009). Online interpretation of scalar quantifiers: insight into the semantics-pragmatics interface. *Cognitive Psychology*, *58*(3), 376–415.

Levinson, S. C. (2000). *Presumptive Meanings - The Theory of Generalized Conversational Implicature*. MIT Press.

Van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (in prep). Scalar diversity.

The case of the inconsistent implicature

Chris Cummins (University of Edinburgh)
c.r.cummins@gmail.com

Over the past decade, a considerable volume of experimental work has accumulated on the topic of implicature, and scalar implicature in particular. In some of the simplest cases, researchers have asked experimental participants whether or not specific pragmatic enrichments were available to them: that is, whether they understood "some" to convey "not all", or whether sentences with "some" were compatible with situations in which "all" was the case.

The results of these experiments have generally been interpreted as offering clear support for competing theories: for instance, for a traditional Gricean approach to implicature (see for instance Geurts 2010) or for a default (Levinson 2000) or grammatical (Chierchia 2006) account. However, on closer inspection, the data have often defied easy characterisation, with implicatures apparently being endorsed at widely varying rates depending upon the precise nature of the task and materials.

For instance, a heated recent debate has concerned the availability of so-called "embedded implicatures". Geurts and Pouscoulous (2009) argue that these are predicted to arise freely under the grammatical account of SIs, but show that they are in practice often unavailable. In response, Chemla and Spector (2011) contend that the sometime availability of embedded SIs – which they demonstrate in a further experiment – is inexplicable within a classically Gricean approach. The discussion continues.

Here I would like to focus attention on the apparent inconsistency, or lack of unanimity, in the behaviour of participants, even within the same experimental paradigm. Put crudely, this has resulted in a situation in which something occurring 30% of the time is interpreted as evidence for it being impossible, and something occurring 70% of the time as evidence for it being obligatory. I would argue that we need to be able to characterise the behaviour of individual participants more precisely if we are to explain the full richness of the data in terms of any theory.

Although I won't be able to explain the observed behaviours in this talk, I will discuss some research that strikes me as offering useful angles of attack. In particular, I look at the implications of "pragmatic tolerance", as argued for by Katsos and Bishop (2011), and consider how this interfaces with cognitive and personality traits (drawing upon recent work with Antoniou and Katsos). And I consider how the recent results of Geurts and colleagues concerning the variability between different implicature triggers might help us understand the processes and responses that we elicit in the lab.

References

Chemla, Emmanuel & Benjamin Spector. 2011. Experimental evidence for embedded scalar implicatures. *Journal of Semantics* 28. 359-400. http://dx.doi.org/10.1093/jos/ffq023.

Chierchia, Gennaro. 2006. Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry* 37. 535-590. http://dx.doi.org/10.1162/ling.2006.37.4.535.

Geurts, Bart & Nausicaa Pouscoulous. 2009. Embedded implicatures?!? *Semantics and Pragmatics* 2. 1-34. http://dx.doi.org/10.3765/sp.2.4.

Geurts, Bart. 2010. *Quantity Implicatures*. Cambridge: Cambridge University Press.

Katsos, Napoleon & Dorothy V. M. Bishop. 2011. Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition* 120. 67-81. http://dx.doi.org/10.1016/j.cognition.2011.02.015

Levinson, Stephen C. 2000. *Presumptive Meanings*. Cambridge, MA: MIT Press.

"Not all" comes for free

Francesca Foppolo (University of Milano-Bicocca) and Marco Marelli (University of Trento)
francesca.foppolo@unimib.it

Controversial evidence is shown as for the incremental processing of scalar terms like "some": Huang & Snedeker (2009) found that "some" requires additional processing time than "all", while no delay was found by Grodner et al. (2010) and Degen & Tanenhaus, 2013.

Our study. Using a Visual-World paradigm we tested Italian V(O)S sentences like "Adesso invece fanno il poliziotto alcuni dei/tutti i pentagoni/triangoli" [lit:*now are policeman some of/all of the pentagons/triangles*] in a situation in which all the pentagons (but only some of the triangles) are policemen (Fig.1) in 3 conditions: underinformative-some (some-pentagons-are-policeman) that is logically true but pragmatically false by virtue of the Scalar Implicature "some but not all"; all-true (all-pentagons-are-policeman); all-false (all-triangles-are-policeman). A truth-value judgment was required and anticipatory eye movements were measured at the onset of the quantifier towards the notALL target (the triangles) compared to the ALL target (the pentagons). The display arrangement and the sentential word order allowed for early and rapid anticipation at the quantifier region: differently from previous study, the quantifier followed the predicate and the contrast set for deriving the some-but-not-all inference was self evident within a single quadrant, and did not require to consider the other quadrants nor the initial object distribution. **Results.** Of the 22 participants, 16 mostly rejected (>75%) underinformative-some (pragmatic responders); 5 mostly accepted them (logic responders) and 1 provided inconsistent answers (excluded). Eye movement data were submitted to a series of logistic regression models (Jaeger, 2008) in which the likelihood of looking at the target quadrant notALL vs. ALL (e.g. triangles/pentagons) was modelled as a function of the experimental condition in different time regions (predicate/quantifier/post-verbal NP). A statistical difference is found in the quantifier region in pragmatic-responders for "some" (solid-red line, Fig.2) compared to the other conditions ($z = 2.387$, $p = .0171$): pragmatic-responders looked more to the notALL quadrant when hearing "some". RT analyses did not show any difference bw. responder types: with a power level of > .99 (even when conservatively assuming extremely small effect sizes (d=.01), Cohen, 1988) we can safely reject the alternative explanation that pragmatic responders are slower than logic responders. **Conclusions**. We show evidence for rapid, cost-free integration of pragmatic-some. We argue that this is due to the fact that the enriched some-but-not-all interpretation was made all the more salient in our study, being the scalar quantifier in a focus position, and being the contrast set for interpreting some pragmatically within the same quadrant (some-but-not-all the triangles), making the scalar alternatives readily available for interpretation. Theoretical and methodological implications of our findings will be discussed.

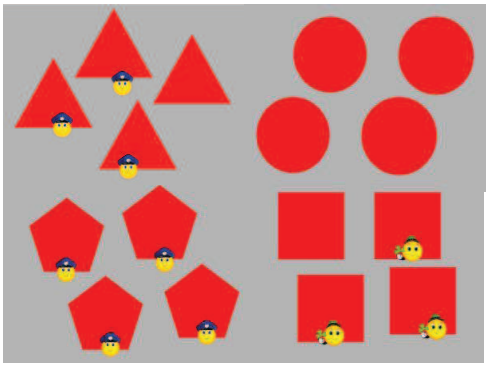


Fig 1 *now are policemen some of/all the pentagons/triangles*

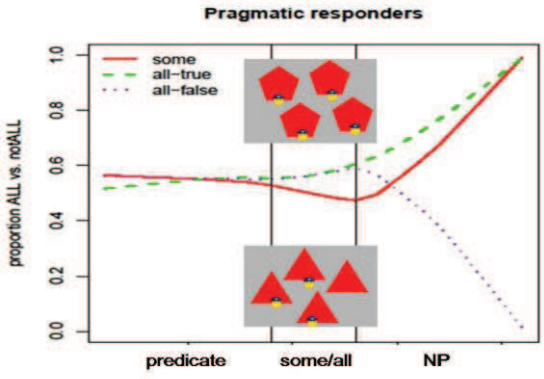


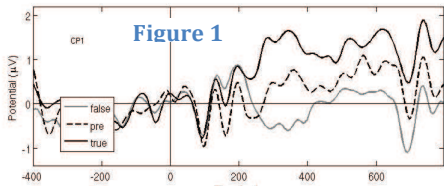
Fig. 2 Eye-movement pattern of pragmatic responders across conditions (some/all-true/all-false)

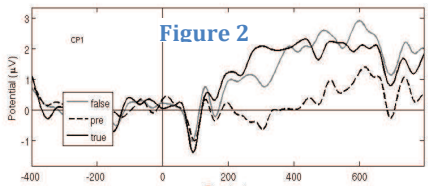
Some implicatures take their time. An ERP study on scalar implicatures with 'sentence-picture vs. picture-sentence' verification task.

Daniele Panizza & Edgar Onea (University of Göttingen).

A scalar implicature is the mechanism through which we arrive at the pragmatic interpretation of *some* ('some but not all'). Under derivational accounts (cf. [1]) this process requires prior access to semantic interpretation ('some and maybe all'). Instead, under probabilistic theories (cf. [2,3]) both meanings are immediately available and the implicature is the process by which one or the other meaning is selected. Growing experimental evidence shows that the pragmatic meaning is more costly than the semantic one [4]. While this follows from derivational theories, it is accounted for by probabilistic theories by assuming that this interpretation is inherently more complex and thus it imposes processing effort either during encoding or verification phase. To tackle this fundamental issue we designed an ERP experiment employing a sentence-picture verification task in which we manipulated a) the quantifier (*all* vs. *some*), b) the match between the sentence interpretation and the visual scenario and c) the order of presentation of sentence vs. pictures. The experiment was conducted in German (*einige* = some, *alle* = all) with 24 participants. The design is described as follows: each trial showed an initial scenario with a character (e.g. the father) who intends to perform an action involving change of state of four objects of the same kind (e.g. 'eat apples'). Familiarity with the objects was achieved by showing them in the introductory slide in the initial and final state. Then the four pictures depicting what really happened were displayed, either followed or preceded by the target sentence. The target sentence (e.g. "He ate some/all of the apples") was uttered by another fictional character (Uncle Peter) and the participant had to judge at the end of the trial whether he is right or wrong. The sentence and pictures were presented at the center of the screen word by word and token by token, respectively. The four pictures could appear in three combinations (e.g. A ='apple' vs. E='eaten apple': EEEE, EEEA and EEAA). The first combination is true for *all* and pragmatically false for *some* whereas the second and third are true for *some* and false for *all*.

The manipulation of the order of presentation allows a comparison between the encoding of quantifier meaning with or without pictorial support. Both quantifiers exhibited a (N400-like) sustained negativity, maximum on centro-parietal sites, from 200 ms when presented before the pictures. However, with *all* the prior presentation of the pictures which did not match its meaning led to greater negativity in the same time-window vs. smaller negativity for matching pictures (Fig.1). This suggests that meaning encoding process and semantic violation affect the same brain network. Crucially, with *some*, encoding-related negativity was elicited by the sentence-first vs. picture-first condition (Fig.2) but no such effect came about in the comparison between pragmatic violation and true condition for *some* (contra the findings of [5]). The effect of the pragmatic violation, instead, showed up at 600 ms as a negative wave on left-anterior electrodes (L-LAN), an effect found in other studies investigating logical contradictions [6,7]. This supports the derivational accounts according to which pragmatic enrichment must follow the access to the semantic meaning. The sentence-first condition did not display any early negativity effect on the critical picture but a frontal positivity from 500 ms (FP600), which has been dubbed as correlate of discourse restructuring [8], for pragmatically false *some* compared to true *all*. This suggests that a scalar implicature was already added to the meaning of the sentence before the pictures have been presented. Our final conclusion is that the complexity of pragmatic readings of *some* affects both the encoding and the verification processes at a relatively late stage, after that initial semantic processing is performed.





References: 1. Huang & Snedeker (2008) Cog Psy 2. Grodner et al. (2010) Cognition 3. Degen (2009) thesis 4. Bott et al.(2012) JML 5. Politzer-Ahles (2013) et al., Brain Res 6. Shao & Neville (1991) 7. Steinhauer et al. (2010) Neuropsy. 8. Kaan & Swaab (2003) J Cog Neuro

The local contrast expectation in *let alone* coordination

Jesse A. Harris (Pomona College) & Katy Carlson (Morehead State University)
Jesse.Harris@pomona.edu

Natural language possesses numerous coordinators that impose different structural and pragmatic constraints on their conjuncts. Here, we explore the unique constraints on *let alone* coordination (LAC), e.g, *John didn't run a MILE, let alone TWO* (Fillmore et al., 1988). Although *let alone* appears to combine with almost any type of syntactic phrase and can occur in various positions within a clause, e.g., *John didn't RUN, let alone SPRINT, a mile*, it always involves gapping of the second conjunct (Hulsey, 2008; Harris, 2013). We find a strong preference for LAC to contrast with local constituents (1), as well as a preference for the *let alone* phrase to occur after a complete clause (2) which ultimately follows from the processor's expectation for default focus at the end of a clause, as with ellipsis generally (e.g., Carlson et al. 2009).

Evidence for (1-2) was detected in a study of the British National Corpus (BNC), in which LAC typically operated over NP (47%) or VP (37%) conjuncts. In 1235 instances of LAC in the BNC, over 80% followed a complete clause, supporting (2). Within the NP set, LAC predominantly contrasted with objects (*Clinton won't withdraw the troops, let alone special forces*) over subjects, which were only 17% of the sample (*Clinton won't withdraw the troops, let alone McCain*). VP instances also showed a locality preference (76% local contrasts). A ratings study (N=20; 16 items) tested (1) and (2) by varying the placement of LAC in sentences like (3). Interrupting a clause with LAC was rated less natural than when it appeared after a clause boundary, and this penalty was increased when the contrast was not local. Together, the results suggest that, though LAC is syntactically flexible, individuals should expect LAC to contrast with a local constituent while respecting clause boundaries.

We further explored whether violating the expectation of a local contrast (1) would be evident during online processing in two self-paced reading experiments. The first study (N=16; 16 items) placed a contrastive adjective (*nicest*) in either the subject or object phrase (3), which formed a natural contrast with an adjective in the LAC conjunct. In region 5 (the second conjunct), there was a 137ms cost for subject over object contrasts, t=2.90, which diminished over time. A second study (N=32; 16 items) set up expectations for contrast through a preceding question, as in (4), which manipulated the placement of coordination (subject or object). When the LAC contrasted with the subject, violating the expectation for a local (object) contrast, there was an immediate 76ms increase in reading time on the second conjunct, t=2.24. These results support the idea that comprehenders expect LAC to contrast with the most local constituent and use semantic and discourse information to determine which elements are in contrast. Though immediate, such information does not overturn the expectation for local contrast.

Most of the research on structural expectations for coordinate structure has explored a small subset of coordinators, primarily *and* and *but* (e.g., Frazier 1987; Hoeks et al., 2006). Investigation of other coordination types as above reveals that they pattern with ellipsis generally in preferring to contrast with later arguments, most likely because these constituents are often focused. Further, we find evidence of two potentially competing constraints on contrast positions, while enriching our understanding of how the processor uses syntactic and discourse expectations when processing coordination.

(1) **Locality constraint**: Preferentially contrast with the most local constituent possible.
(2) **Clause-final preference:** Don't interrupt a clause with *let alone* coordination.
(3) $|_1$The (nicest) nurse $|_2$ couldn't stand $|_3$the (nicest) patient, $|_4$ let alone $|_5$ the meanest one, $|_6$ and no one at the hospital $|_7$ was happy at all.
(4) *Subject:* Did Mary and Diana call John? / *Object:* Did Mary call John and Diana?
$|_1$ Mary didn't call John,$|_2$ let alone Diana, $|_3$ and I'm very upset $|_4$ about it all.

Language affects Quantity Judgments in bilingual Yudja speakers

Suzi Lima (Harvard University)
lima@fas.harvard.edu

This paper contributes to investigations of whether language-specific encoding of the count mass distinction affects the construal of entities (cf. Imai and Getner 2007, Li, Duham, Carey 2009). Two quantity judgment studies tested adult bilingual speakers in two languages that encode the count-mass distinction differently: Yudja (an indigenous language spoken in Brazil by 348 people) and Brazilian Portuguese (BP). In Yudja all nouns have count denotations. That is, all nouns (including notional mass nouns like 'water') can be directly combined with numerals and with count quantifiers such as 'many'. Conversely, in BP count and mass nouns are grammaticalized in different ways. For example, only count nouns can be directly combined with numerals and count quantifiers. In two studies with 20 Yudja bilingual adults we investigate whether the construal of entities of a bilingual speaker can change depending on the language that he/she is operating.

Quantity judgments task Materials and methods (Barner and Snedeker (2005) paradigm) While presenting two different drawings, one with a big portion of *x* (Volume drawing) and another with three different portions of *x* (Number drawing), we asked: *Ma de bitu x dju au?* ('Who has more *x* ?'). Subjects answered 3 questions with a notional mass noun (e.g., *asa* 'flour'), 3 questions with a notional count noun (e.g., *xãã* 'bowl') and 2 questions with an aggregate noun (e.g., *abeata* 'clothes'). Participants had to point to one of the drawings to answer the question. Results Yudja speakers consistently chose the 'Number' drawing for all noun categories (notional mass nouns: 85% of 'Number' responses; notional count nouns: 83% of 'Number responses'; aggregate nouns: 79% of 'Number' responses). Mixed effects modeling using Helmert contrasts confirmed that there was no effect of noun type. Conclusion The default reading for notional mass nouns like *water* in Yudja is not a mass reading, but a count reading (concrete portions of *x,* e.g., piles of flour).

Quantity judgments tasks in BP Materials and methods (Barner and Snedeker (2005) paradigm) While presenting two different drawings, one with a big portion of *x* (Volume drawing) and another with three different portions of *x* (Number drawing), we asked: *Quem tem mais x?* ('Who has more *x* ?'). Subjects answered 3 questions with a notional mass noun (e.g., *farinha* 'flour'), 3 questions with a notional count noun (e.g., *cuia* 'bowl') and 2 questions with an aggregate noun (e.g., *roupa* 'clothes'). Participants had to point to one of the drawings to answer the question. Results Mixed effects modeling using Helmert contrasts confirmed that there was an effect of noun type when we contrasted count and mass nouns in BP. It was observed that mass nouns are significantly less likely to be associated with 'Number' responses in comparison to count nouns (Wald's $Z = -2.256$; $p = 0.02408$; $\beta = -0.48$). There was also a significant effect of Age in proportion of 'Number' responses as younger bilingual speakers tend to differentiate count from mass nouns in most trials in contrast to older bilingual speakers (Wald's $Z = -2.33$; $p = 0.019$; $\beta = -0.21$). **Conclusion** The default reading for mass nouns like 'water' in BP is a mass reading, while for count nouns and aggregates it is a count reading.

General discussion the results show that Yudja bilingual adults have different quantity judgments in Yudja (L1) and BP (L2). In BP, a language that grammatically distinguishes count from mass nouns, there was a significant effect of noun type, which was not observed in Yudja, a language where all nouns have count denotations. The results of these studies suggest that a single individual's 'construal of the world' *can* change as one changes the language that the individual is operating in.

References Barner, D., & Snedeker, J. (2005). Quantity judgments and individuation: Evidence that mass nouns count. *Cognition*.

Imai, M. & Mazuka, R. (2007). Revisiting language universals and linguistic relativity: language- relative construal of individuation constrained by universal ontology. Cognitive Science, 31, 385- 414.

Li, P. Dunham, Y., Carey, S. (2009). Of Substance: The Nature of Language Effects on Entity Construal. In: *Cognitive Psychology*. June; 58(4): 487–524.

Time heals semantic illusions, but not syntactic illusions

Dan Parker & Colin Phillips (University of Maryland) & Alan Du (Montgomery Blair High School)
dparker3@umd.edu

Illusory licensing effects have been argued to provide a window into how information in recent memory is accessed during comprehension, suggesting the use of a direct access mechanism in which structural information has no privileged status. In previous research, there has been a consensus that the time profile of the illusory licensing effects varies as a function of the amount of time after the introduction of the licensee. These accounts predict that the representation of existing structure is stable over time, and that variation in the time or position where the licensee is introduced in the sentence should not impact the illusion. We present results from self-paced reading (SPR) and speeded acceptability judgments (SAJ) that show that one type of illusion, which depends on semantic/pragmatic licensing mechanisms, shows a fleeting time profile, such that it is present or absent depending on the timing of when the licensee is introduced. But we find that the same time profile does not extend to another type of illusion, which depends on a morphosyntactic licensing mechanism. We argue that the contrast reflects changes over time in the encoding of emerging semantic/pragmatic representations, and we provide computational modeling evidence in support of these findings.

Illusory licensing is observed for negative polarity items (NPIs) and agreement, manifesting as eased processing of illicit licensees due to a structurally irrelevant licensor (1). Although NPIs and agreement are licensed using different mechanisms, e.g., semantic/pragmatic licensing vs. morphosyntactic licensing, illusory licensing effects in both cases suggest that the parser can access semantic and morphosyntactic features independently from structural features online.

(1) a. *The candidates [that **no democrats** disliked] have **ever** won. (*NPI 'ever' illusion*)
b. *The key to **the cabinets** apparently **were** destroyed in the fire. (*agreement illusion)*

Expts. 1a-b (SPR, SAJ, n=24, Parker et al. CUNY 2013) tested the prediction that the time or position of where the NPI is introduced should not impact the illusion. Results disconfirmed this prediction, as main clause NPIs showed the illusion, but embedded clause NPIs did not (2). This contrast could reflect the structural position of the NPI or the time between the licensors and the NPI. Expt. 2 (SAJ, n=24) distinguished these alternatives by holding constant the structural position of the NPI and manipulating the position of a parenthetical phrase to vary the time between the licensors and the NPI (3): the illusion disappeared for sentences involving an intervening parenthetical phrase. Expt. 3 (SAJ, n=24) tested whether the contrasts observed for NPIs extends to agreement using a similar parenthetical phrase manipulation, (4): the position of the parenthetical phrase did not modulate the agreement illusion.

(2) The journalists [that **no editors** recommended] (**ever**) thought that readers would (**ever**) understand physics.
(3) (As the secretary noted), the journalists [that **no editors** recommended] (as the secretary noted) **ever** received compensation.
(4) (As the janitor mentioned), the key **to the cabinets** (as the janitor mentioned) probably **were** destroyed in the fire.

NPI illusions, but not agreement illusions, show a fleeting time profile, and the presence or absence of the NPI illusion depends on the timing of when the NPI is introduced. This contrast supports the claim that these dependencies are qualitatively different, but the fleeting profile for NPI illusions is not predicted by existing accounts. We argue that the contrast reflects changes over time in the encoding of emerging semantic representations, such that semantic licensing features are no longer independently accessible, preventing further partial-match illusions. In support of this contrast, we provide simulations in which the feature-bindings in a representation change over time from a transparent to an opaque encoding.

The relationship between regressive saccades and the P600 effect: Evidence from concurrent eye movement and EEG recordings

Paul Metzner (University of Potsdam), Titus von der Malsburg (St John's College, University of Oxford), Shravan Vasishth (University of Potsdam), Frank Rösler (University of Hamburg)
paul.metzner@uni-potsdam.de

Reading studies using eyetracking have found that word integration difficulties are reflected in elevated reading times and regressive eye movements (Clifton, Staub, & Rayner, 2007). Studies measuring event-related brain potentials (ERPs) have found that semantic and syntactic integration difficulties also elicit characteristic brain responses: N400 effects for semantic (Kutas & Hillyard, 1980) and P600 effects for syntactic integration (Osterhout & Holcomb, 1992). For two reasons, it is not clear how eyetracking and ERP effects map onto each other. First, no regressions occur in ERP designs where sentences are presented word-by-word (RSVP). Second, eye movements in natural reading produce artifacts in the EEG, which until recently complicated the analysis considerably. Thus, the precise relation of eye movements and ERPs is unclear.

To address this issue, we conducted a study of both types of integration difficulty in which we recorded eye movements and brain potentials concurrently and used ICA (Jung et al., 2000) to correct the eye movement artifacts in the EEG. The materials in our study followed a design developed by Hagoort (2003):

(1) Der/**Die** verfallene/*neugierige* Bauernhof braucht eine Renovierung. Er ist…
The$_{[masc]}$/**The**$_{[fem]}$ run-down/*inquisitive* farm needs a renovation. It is…

(2) Der erfahrene Star spielt die/**das** schwierige/*elektrische* Rolle. Er überzeugt…
The experienced star plays the$_{[fem]}$/**the**$_{[neut]}$ difficult/*electric* role. He convinces…

Sentence-internally (1) or sentence-finally (2), sentences could contain a gender mismatch between determiner and noun (syntactic violation, bold face for violating determiner, critical noun underlined) or an adjective that did not match with the noun semantically (semantic violation, italics). To have a baseline for the natural reading ERPs, we presented the same sentences using RSVP and randomly assigned participants (N=72) to one of the two presentation modes.

We replicated the results of the Hagoort study with RSVP (N=24): Syntactic violations elicited a P600 in sentence-internal position and an N400-P600 response in sentence-final position, semantic violations an N400-P600 in sentence-internal and an N400 in sentence-final position. The ERPs obtained from natural reading (N=48) were qualitatively similar when we analyzed the entire data set. When we analyzed only violation trials with regression from the critical noun, we found a P600 in sentence-internal position and an N400-P600 in sentence-final position for syntactic and semantic violations. In contrast, we found no P600 in either position and for either syntactic or semantic violations when no regression occurred. In sentence-final position, we instead found a sustained centro-parietal negativity for both violations.

This demonstrates two things: First, the ERP results from natural reading replicate the results from RSVP, which is further evidence for the method's validity (see also Dimigen et al., 2011; Kretzschmar et al., 2009). Second, the results show that the P600 effect is associated with regressions in natural reading (see also Dimigen et al., 2007), lending support to the idea that recovery processes are accompanied by a P600 effect. The fact that such an effect was absent in trials without regression suggests that the parsing system pursued a qualitatively different processing strategy that may not involve attempted recovery. Thus, the benefit of the coregistration approach is two-fold: First, it provides us with a behavioral signal that enables us to better understand ERPs. Second, the ERP results allow us to attach functional interpretations to eye movement phenomena like regressions.

Predictability and gap-filler ordering in dependency formation: An MEG Study

Kimberly Leiken & Liina Pylkkänen (New York University)
kal412@nyu.edu

Introduction: In the neurolinguistic literature, an increase of hemodynamic activity in the left inferior frontal gyrus (LIFG, aka "Broca's Area") is found for ORs (which contain a filler item long before its associated gap) compared to subject relative clauses (SRs) (e.g. Caplan et al., 2000, 2008; Constable et al., 2004; Just et al., 1996; Keller et al., 2001; Stromswold et al., 1996), mirroring the behavioral finding that ORs are more costly than SRs. Although these results originally led to the conception that Broca's area is linked to syntactic processing, more recent behavioral studies suggest that ORs are only costly when involving stimulus material eliciting similarity-based interference (Lewis, 1996; Gordon, Hendrick, & Johnson, 2001; Gordon, Hendrick, & Levine, 2002; Lee et al., 2007, Van Dyke & McElree, 2006). Recently, it has been shown that the LIFG increase is also dependent on such interference, i.e., no effects were obtained for the sheer presence of a dependency (Leiken, K. & Pylkkänen, L., 2013). This suggests the bulk of dependency processing may occur earlier, upon encountering the filler item. In contrast with ORs, which always involve a cue indicating an upcoming dependency formation, the present study investigates dependencies that do not have this cue. Verb Phrase Ellipsis (VPE) was used as an example of a non-predictable dependency. Without an early indication of a dependency, we might expect the bulk of neural activity to take place at the ellipsis site. Additionally, we included a construction in which the so-called "filler-gap" order is reverse to that of ORs, namely Right Node Raising (RNR). This order reversal begs the question of whether we might still expect an early LIFG effect of dependency, or whether this reversal might delay dependency processing. Using magnetoencephalography (MEG) we compared neural activity in response to ORs, VPE constructions, which are not well understood, and RNR constructions, which have never before been studied in the imaging literature, to determine whether the established LIFG effect might be modulated by predictability or gap-filler ordering.

Design: VPEs were specifically selected because they contain a dependency that cannot be pre-processed. RNRs were selected for their reversal of the OR filler-gap order. Obligatorily transitive verbs were used in the 1st clause of all conditions to prevent interpretation of RNR-Test as a basic conjunction. "*Did*" (which has an auxiliary verb use) was not used in the 2nd conjunct of ORs/RNR to prevent a VPE reading.

VPE-Test	The husband hogged	the blankets	and Jane	did		too.
VPE-Control	The husband hogged	the blankets	and Jane	did	that	too.
OR-Test	The husband hogged	the blankets	that Jane	grabbed		afterward.
OR-Control	The husband hogged	the blankets	and Jane	grabbed	them	afterward.
RNR-Test	The husband hogged		and Jane	grabbed	the pillows.	
RNR-Control	The husband hogged	the sheets	and Jane	grabbed	the pillows.	
Filler	The husband was in	the kitchen	and Jane	did	laundry.	

The target regions of analysis were at the site of dependency formation; at the gap in ORs, at the filler item in RNR, and at the ellipsis site in VPE. Non-parallel versions of each condition (which contained a proper name in place of the interfering Determiner-Noun Phrase, and did not contain an internal RC) were included.

Hypotheses: If ORs allow for early dependency processing at the filler item, unpredictable VPE dependencies should cause a delay in LIFG activity until immediately after the site of the ellipsis. If the gap in RNR behaves similarly to the filler in ORs, as an indicator of future dependency formation, we would expect no LIFG activity at the RNR filler site. However, if the gap alone fails to signal an upcoming dependency, then we would expect to find an effect in the LIFG at the filler site.

Results: Consistent with previous results, no LIFG effect of pure dependency was found at the gap site in ORs. Nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007) on LIFG activity in VPE identified a significant main effect of dependency formation at 200-344ms (p=.0284), suggesting that LIFG activity increases at a retrieval site when the dependency cannot be predicted. A significant main effect of dependency was also found in RNR at 200-400ms (p=.0058), indicating that despite the potential for anticipatory dependency formation, the gap-first ordering in RNR forces a delay in dependency processing. This research unveils a complex pattern of interactions between dependency formation and similarity-based interference unfolding at different time windows.

Facilitatory semantic interference in the processing of long distance dependencies

Ian Cunnings (University of Reading) & Patrick Sturt (University of Edinburgh)
i.cunnings@reading.ac.uk

Similarity-based interference has been identified as a primary factor in the resolution of linguistic dependencies [4]. In reading studies, such interference often incurs longer reading times at a point of retrieval when a distractor matches the content of the retrieval target in some way (e.g. animacy).The interpretation of such inhibitory interference is however debated. Some claim that such effects index retrieval of the distractor [2,3]. Others have argued that inhibitory interference need not implicate distractor retrieval, but instead may index retrieval of a degraded target representation [1]. Facilitatory interference on the other hand, as is found in agreement attraction, provides stronger evidence of the retrieval of a distractor [1,5]. Using a plausibility diagnostic manipulated after the point of retrieval, we aimed to tease apart these two accounts by examining whether inhibitory interference facilitates subsequent semantic interpretation.

32 native English speakers read 24 critical items as in (1), plus 104 fillers, while their eye-movements were monitored. In (1a/b), *the policeman* is a plausible agent of 'arresting' whereas *the criminal* in (1c/d) is implausible. The grammatically inaccessible distractor (*cop/car*) is animate in (1a/c) and inanimate in (1b/d). Inhibitory interference is expected at the retrieval site 'shouted'. If this effect does not index retrieval of the distractor, a main effect of plausibility is expected at or after 'arresting the robber'. Evidence of attenuation of the plausibility effect in (1c), when the distractor is a plausible agent of 'arresting', would provide evidence that properties of the distractor are indeed sometimes retrieved.

(1a) A bad crime had been committed. The policeman who saw that the cop was waiting shouted loudly after arresting the robber for theft.
(1b) A bad crime had been committed. The policeman who saw that the car was waiting shouted loudly after arresting the robber for theft.
(1c) A bad crime had been committed. The criminal who saw that the cop was waiting shouted loudly after arresting the robber for theft.
(1d) A bad crime had been committed. The criminal who saw that the car was waiting shouted loudly after arresting the robber for theft.

At the retrieval site (*shouted*) reading times were reliably longer in (1a/c) than (1b/d), indicating inhibitory similarity-based interference, in second pass times ($p = .024$). Inhibitory interference was also observed at the spillover region (*loudly after*). At the critical region (*arresting the robber*) we observed reliable main effects of plausibility that were modulated by reliable plausibility by distractor interactions in second pass and total reading times ($ps < .046$). In both measures, reading times were generally longer in implausible conditions (1c/d) than (1a/b), but this plausibility effect was reliably attenuated in (1c) compared to (1d) ($ps < .049$). At the final region (*for theft*), reliable main effects of plausibility, not modulated by the distractor, were observed in regression path times, second pass times and total reading times ($ps < .025$).

To our knowledge, this is the first study to report facilitatory similarity-based semantic interference during online semantic interpretation for long-distance dependencies. That we observed interactions at the critical region as well as main effects of plausibility at the sentence final region is similar to the interference effect observed for agreement attraction [1,5]. Together, these results provide strong evidence that properties of a grammatically inaccessible distractor constituent are sometimes retrieved during the resolution of long distance dependencies.

References
[1] Dillon et al. (2013), JML 69: 85-103. [2] Van Dyke (2007), JEP: LMC 33: 407-430. [3] Van Dyke & McElree (2011), JML 65: 247-263. [4] Van Dyke & Johns (2013), LLC 6: 193-211. [5] Wagers et al. (2009), JML 61: 206-237.

Demonstrative equatives and the conveyance of speaker (un)certainty

Gregory Ward (Northwestern University)
gw@northwestern.edu

In this talk (representing collaborative work), I analyze two English copular constructions, both with a demonstrative pronoun in subject position: epistemic *would* equatives (*That would be John*) and *that*-equatives (*That's John*). While the syntax and semantics of these constructions is a matter of intense ongoing debate (Hedberg 2000; Heller & Wolter 2008; Mikkelsen 2007; *inter alia*), their information structure has received far less attention (cf. Birner, Kaplan, & Ward 2007).

Drawing upon a large corpus of naturalistic data, I show that the modal in an epistemic *would* equative serves to mark the FOCUS of the utterance, thus requiring that an OPEN PROPOSITION (Prince 1986) be contextually salient, with the post-copular constituent serving as the instantiation of the variable of that open proposition (OP). The information structure of the epistemic *would* construction accounts for the humorous and/or ironic tone often associated with its use. The *that*-equative construction, on the other hand, is more pragmatically constrained. It may also be used to instantiate an OP; however, for *that*-equatives, unlike epistemic *would* equatives, such a possibility is determined contextually rather than morpho-syntactically.

As for the interpretation of the two constructions, I present the results of a series of empirical studies that show that use of an epistemic *would* equative conveys a high degree of speaker commitment to the truth of the expressed proposition. In our first perception study, participants read short conversations and rated inferred speaker certainty of target sentences with or without epistemic *would*. This textual condition allowed us to study the effect of modality on the assessment of speaker certainty, in isolation from the effect of intonation. We found that modal *would* alone was rated no more uncertain than main verb *be*; a one-way ANOVA reported no significant difference between the two mean ratings of certainty ($F(1, 238) = 0.58$, $p = 0.45$).

In our second perception study, participants were presented with a series of contextually-appropriate epistemic *would* and *that*-equatives produced with three different intonational contours, and again asked to rate inferred speaker certainty associated with the target construction. This spoken condition allowed us to study the effect of both epistemic modality and intonational contour, combined or in isolation, on the assessment of speaker certainty. We found that use of epistemic *would* equatives (observed mean = 0.15) tended to be perceived as significantly more certain than *that*-equatives (mean = 0.01). For contour, we found that downstepped utterances were perceived significantly as most certain (mean = 0.7), followed by declarative utterances (mean = 0.48); not surprisingly, the *yes-no* question contour was perceived as least certain (mean = –0.93). A two-way ANOVA revealed no significant interaction between type of equative and type of intonational contour ($F(2, 594) = 0.698$, $p = 0.498$).

From these findings, we conclude that, far from being a marker of tentativeness as has been claimed (Palmer 1990, Perkins 1983), use of epistemic *would* (with any of three distinct intonational contours) conveys an even higher degree of speaker certainty than does use of the corresponding *that*-equative produced with the same contour.

References

Birner, B., Kaplan, J., and Ward, G. (2007). Functional compositionality and the interaction of discourse constraints. *Language* 83:323:349.

Hedberg, N. (2000). The referential status of clefts. *Language* 76:891-920.

Heller, D. and L. Wolter. (2008). That is Rosa: Identificational sentence as intentional predication. In A. Grønn, (ed.), *Proceedings of Sinn und Bedeutung* 12:226-240. Oslo: Dept. of Literature and Languages, University of Oslo.

Mikkelsen, L. (2007). On so-called truncated clefts. In L. Geist and B. Rothstein (eds.), *Kopulaverben und Kopulasatze: Intersprachliche und Intrasprachliche Aspekte,* 47-68. Tübingen: Niemeyer Verlag.

Palmer, F. (1990). *Modality and the English Modals*. New York: Longman.

Perkins, M. (1983). *Modal Expressions in English*. London: Frances Pinter.

Prince, E. (1986). On the syntactic marking of presupposed open propositions. CLS 22:208-222.

Hearing a Who: Preschoolers and Adults Process Language Talker-Contingently

Sarah Creel (UCSD)
creel@cogsci.ucsd.edu

Listeners process sentences, but they also process people. Research in the past few decades indicate that a talker's identity or (perceived) social group influences language processing at a variety of levels: phonological (e.g. Niedzielski, 1999), lexical (e.g. Goldinger, 1996), syntactic (Kamide, 2012), and discourse (Horton & Gerrig, 2005) levels.

Do these instances of talker specificity reflect small-scale flexibility of highly abstract language knowledge, or do they represent a crucial facet of language processing? I argue the latter. At least two critical elements of language processing are profoundly affected by talker identity. First is phonemic category extraction: listeners who are new to a language have difficulty generalizing speech sound and word recognition to new voices, and are aided by voice variability during learning (e.g. L1: Houston & Jusczyk, 2000; L2: Lively et al., 1993). Second are higher-level expectation effects in language processing, at the level of discourse processing and "talker-semantic" encoding. I will touch briefly on issues of phonemic category extraction and word encoding, but I will primarily discuss discourse and semantic aspects of talker identity, including my own research on the development of talker processing.

A variety of studies suggest that language is a powerful cue to social groups (Eckert, 2008). Knowing someone's social group, or even their particular identity, influences on-line sentence processing. Adults in an ERP paradigm who heard identical sentences spoken either by a congruous or incongruous talker (e.g. adult vs. child saying "I want to drink the wine") showed a larger N400 semantic mismatch negativity to the target word when the incongruous talker spoke the sentence (Van Berkum et al., 2008). In my own research, I have shown that preschool-aged children direct eye movements preferentially to shapes of the talker's favorite color when that individual is talking ("Show me the circle"; Creel, 2012). In collaborative work (Borovsky & Creel, in press), 3-10-year olds, as well as adults, activated long-term knowledge about different individuals (e.g. pirates vs. princesses) based who spoke the sentence. Specifically, participants hearing a pirate say "I want to hold the sword" directed eye movements preferentially to a sword picture prior to word onset, despite the presence of other pirate-related (a ship) and holdable (a wand) pictures. This suggests that children can use voice information to identify individuals and activate knowledge that constrains sentence processing in real time. Finally, a new study in my lab suggests that preschool-aged children concurrently encode novel word-referent mappings and novel person-referent mappings.

The studies reviewed here suggest that listeners' language apprehension is affected in real time by inferences of who is speaking. This is much more consistent with an interactive view of language processing than a modular view. Even quite young children appear to condition or contextualize their language input based upon who is saying it, suggesting that language acquisition itself is talker-contingent.

References

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*, 487–509.

Eckert, P. (2008). Where do ethnolects stop? *International Journal of Bilingualism*, *12*(1-2), 25–42.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1166–1183.

Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language*, *47*, 589–606.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582.

Kamide, Y. (2012). Learning individual talkers' structural preferences. *Cognition*, *124*(1), 66–71.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, *94*(3 Pt 1), 1242–55.

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology*, *18*(1), 62–85.

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–91.

Privileged vs. shared knowledge about object identity in real-time referential processing

Mindaugas Mozuraitis, Craig G. Chambers, & Meredyth Daneman (University of Toronto)
m.mozuraitis@mail.utoronto.ca

Successful communication requires sensitivity to shared vs. privileged knowledge, whose status is inferred from various sources including interlocutors' perspective differences. In most studies, perspective is manipulated by varying interlocutors' awareness of the ***existence*** of entities in a physical context (e.g., selected objects are occluded so as to make their existence unknown to one individual). However, perspectives may also differ in terms of the ***identity*** ascribed to an object. For example, one interlocutor can mistake an object for another based on its outward appearance whereas another interlocutor knows the object's true identity. These cases are important to consider because developmental work shows that egocentric knowledge about object identity is more challenging to overcome than knowledge about object existence. Also, the processing of identity information has important theoretical implications for semantic phenomena including referential opacity and speaker reference vs. truth-theoretic meaning.

The current study explores listeners' ability to manage privileged knowledge about the identity of visually-misleading real objects. Pretest 1 confirmed that our selected misleading objects (e.g., a yo-yo that looks like a baseball) would be misidentified based on their appearance. Pretest 2 confirmed that, when participants were given foreknowledge about the actual identity of these objects, these objects attracted eye fixations like regular phonological competitors (e.g., an obvious yo-yo) when a phonologically-overlapping target object was mentioned in a spoken instruction (e.g., *yogourt*). The materials were then used in two on-line referential communication experiments in which listeners followed instructions from a (confederate) speaker in a different room via video-chat. Expt 1 manipulated (i) competitor type and (ii) the knowledge that could be attributed to the speaker (listeners always knew the actual identity of misleading objects). Competitors were either "regular" (actual yo-yo) or "misleading" (yo-yo that looks like a baseball) phonological competitors for a target ("yogourt"), or were visual control objects (actual baseball). In the **shared-knowledge** condition, the true identity of misleading competitors was revealed to both listener and speaker before each trial while the video-chat system was online. In the **privileged-identity** condition, the identity of misleading competitors was revealed only to listeners before each trial, while the video-chat system was temporarily paused. In the **privileged-existence** condition, competitors were covered from the speaker's perspective (i.e., the standard manipulation in past studies) and the identity of misleading competitors was revealed to listeners, all while videochat was paused.

In the privileged-existence condition, listeners showed little consideration of all competitor types compared to the shared-knowledge condition ($p<.05$). Thus, listeners seem very effective at managing privileged information about object existence (as in some past studies). In contrast, in the privileged-identity condition, listeners strongly considered misleading competitors (in fact, to the same extent as regular competitors, $p>.1$) even though in the former condition listeners should know that the speaker could not plausibly be referring to these objects using the initial sounds in the unfolding target name (e.g., a yo-yo that looks like a baseball could not be referred to as yo-yo without privileged knowledge). Expt 2 examined whether prior discourse improved listeners' ability to bypass privileged knowledge. In this case, the speaker had previously referred to misleading objects using names that reflected the speaker's knowledge state (e.g., "baseball", for yo-yo that looks like a baseball). Strikingly, despite this reinforcing information, listeners still strongly considered privileged knowledge about the identity of misleading competitors as the target name unfolded. Together, the results demonstrate listeners' sustained consideration of privileged knowledge about object ***identity*** (relative to privileged knowledge about object ***existence***)**,** even when the speaker's knowledge is explicitly highlighted by prior discourse.

What do you know? ERP evidence for immediate use of common ground during online reference resolution

Les Sikos, Sam Tomlinson, Lauren Sanchez and Daniel Grodner (Swarthmore College)
lsikos1@swarthmore.edu

The distinction between common ground (mutually known) information and that which is privileged to one conversational participant is fundamental to referential communication (Clark 1996; Stalnaker 1978). However, recent evidence on the time-course of reference resolution is mixed. Some results suggest that addressees quickly use the common/privileged ground distinction to narrow the set of potential referents (Heller, Grodner & Tanenhaus 2008; Hanna, Tanenhaus & Trueswell 2003). Other findings suggest that addressees initially ignore ground information, essentially adopting an "egocentric" perspective (Barr 2008; Keysar et al. 2000). A key result in support of the egocentric view comes from visual-world eye tracking studies: addressees systematically look at privileged ground competitors despite knowing that the speaker cannot see them and hence could not be referring to that object (Barr, in press; Brown-Schmidt 2009). One concern with this conclusion is that people do not just fixate entities that are candidates for referential description, but also entities that are merely semantically related to a target (e.g., increased fixations to a hammer when hearing "nail") (Yee & Sedivy 2006). Thus increased eye-movements to a privileged competitor might not reflect referential ambiguity.

We used event-related potentials (ERP) to assess the time-course of perspective use in reference resolution. Definite descriptions with multiple potential referents (e.g., *David shot at John as he…*) elicit a sustained negative shift of the ERP which begins 300-600 ms after word onset and can persist over a second after disambiguation (Nref effect; Van Berkum et al. 2007). Participants (N=35) were instructed to select a referent from a cartoon display of four animals via prerecorded instructions generated by a speaker who could only see three of the animals. Instructions such as (1) were paired with one of five display types (Fig. 1). Participants previewed each display for 4 seconds then fixated the screen center during the auditory command (this central fixation did not impair response accuracy; M=0.97).

If participants are initially egocentric (i.e., consider the privileged object to be a candidate for reference), both the privileged ground competitor (PGC) and common ground competitor (CGC) conditions should induce referential ambiguity and elicit an Nref effect relative to the no competitor (NoC) condition. If instead participants immediately take the speaker's perspective, only the CGC condition should elicit an Nref compared to the NoC and PGC conditions.

ANOVAs showed that CGC elicited an Nref effect relative to both NoC and PGC; PGC did not differ from NoC. This pattern of results suggests that there is no point at which addressees consider a privileged competitor on equal footing with a common ground target (contra Barr 2008). Thus, these findings are consistent with theories of language processing that allow socio-pragmatic information to rapidly influence online comprehension. Exploratory analyses suggest that individual differences in social aptitude may contribute to these patterns.

(1) "Click on the triceratops with the party hat."

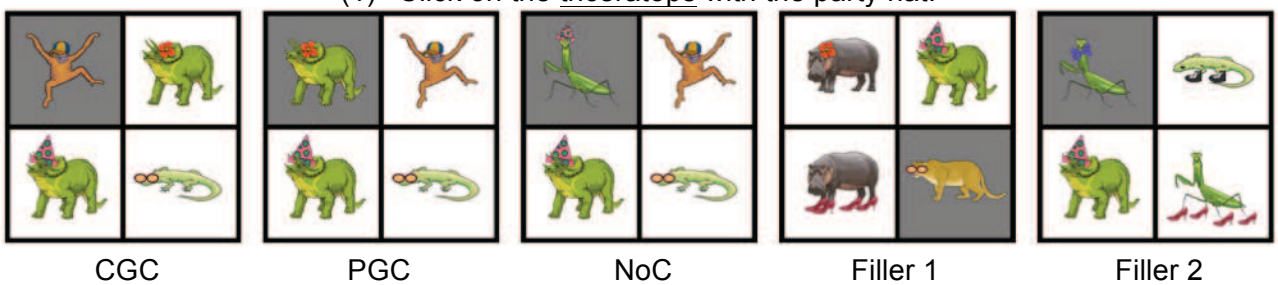


Figure 1. Auditory stimulus (target word underlined) and corresponding participant displays (gray: privileged ground; white: common ground).

A cross-linguistic verb-final bias in gesturing paradigms

Richard Futrell (MIT), Tina Hickey (University College Dublin), Aldrin Lee (University of the Philippines), Eunice Lim (UCSD), Elena Luchkina (Brown University), Edward Gibson (MIT)
Contact: futrell@mit.edu

When asked to communicate simple events such as “boy kicks ball” by gesture, participants create codes that recapitulate the word order patterns of the world’s vocal languages (Goldin-Meadow et al., 2008; Hall et al., 2010; Langus and Nespor, 2010; Gibson et al, 2013; and others). Specifically, they most often convey simple transitive events using gestures in the order Subject-Object-Verb (SOV), the most common word order in human languages. When there is a possibility of confusion between subject and object, participants use the order Subject-Verb-Object (SVO). This overall pattern can be explained by positing a ‘default’ cognitive preference for SOV order, with SVO order emerging to facilitate robust communication in a noisy channel (Gibson et al., 2013).

However, previous studies have not completely examined the possible effects of the gesturer’s native language, in that they have not studied the speakers of typologically diverse languages. Participants in previous studies all spoke either SVO or SOV languages, introducing some possible confounds. For example, SVO and SOV languages might *both* bias one toward subject-initial and verb-final gestures, because both kinds of languages are subject-initial, and intransitive sentences in these languages are verb-final (SV). This kind of issue arises because the space of biases from different kinds of languages has not been adequately explored.

In this paper we demonstrate that there is a verb-final bias in gestures independent of native language structure by eliciting gestures from native speakers of Russian, Irish, and Tagalog, comparing with previous studies of speakers of English, Japanese, and Korean. Irish and Tagalog are VSO languages, allowing us to determine if a verb-final bias exists even for speakers of languages without verb-final constructions.

We find that SOV order emerges in the gestures of speakers of all these languages. The proportions of orders of three-word gestures are shown in Figure 1. The speakers of VSO languages occasionally use verb-initial gestures (17% in Irish and 7% in Tagalog). SOV order is more prevalent for speakers of case-marked languages (in a logistic regression, $p < 0.01$). To gesture ‘reversible’ events (with an animate object), speakers of all the languages primarily switch to SVO gestures (in logistic regressions, object animacy predicts that the object will follow the verb at $p < 0.001$ for all languages).

The findings show that although there is native language bias in gesture order, verb-final order arises regardless of it, and SVO order arises under the same conditions. This provides further support for the idea that improvised gesture is a window into the pressures shaping language formation, independently of the languages that participants already know.

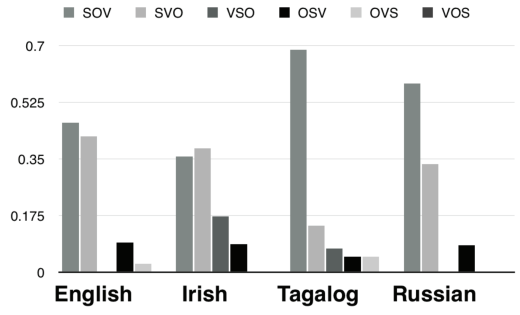


Figure 1. Word orders used in gestures by speakers of four languages, for **non-reversible** events.

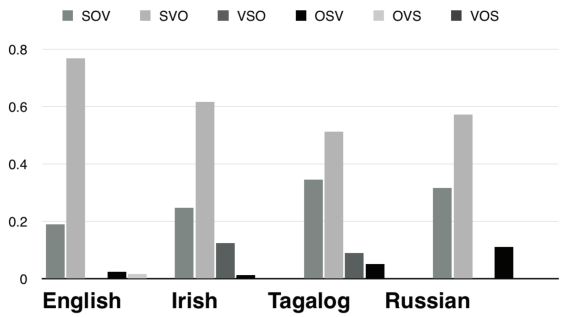


Figure 2. Word orders used in gestures by speakers of four languages, for **reversible** events.

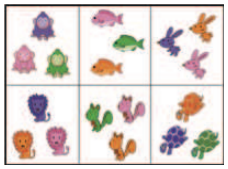
Processing of pitch prominence in Williams syndrome

Kiwako Ito, Marilee Martens, Erin McKenna (Ohio State University)
ito@ling.ohio-state.edu

The highly verbal behavior of individuals with Williams syndrome (WS) has been taken as evidence for the separation of linguistic abilities from general intelligence (e.g., Bellugi et al., 2000). However, recent studies suggest that linguistic skills in WS are at levels predicted by the individuals' mental age (e.g., Brock, 2007). While a small number of studies on prosody in WS also show that comprehension of prosodic cues is delayed in WS compared to chronological age-matched (CA) controls (e.g., Stojanovik, 2010), it is not clear whether the observed lag is due to their general lack of sensitivities to prosodic cues or if it reflects atypical prosodic processing. Using a visual world paradigm that has been shown to capture the immediate responses to prosodic cues (e.g., Weber et al., 2006; Ito et al., 2012), the present study investigates the online responses to contrast-marking prosody in WS.

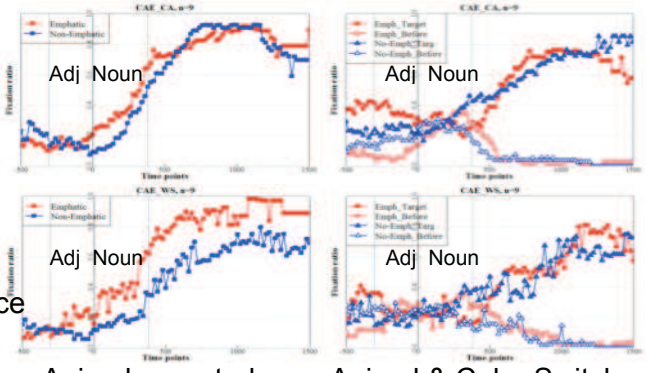
Participants saw slides with animal drawings and clicked on the inquired targets following the pre-recorded auditory stimuli. In each of 50 (2 practice, 24 critical, 24 filler) trials, participants saw 18 animals divided into 6 cells (Example Slide) and clicked on two animals following a paired question. Half of the critical trials repeated the animal (Where is the pink lion → Now, where is the orange lion?) while the other half switched both the color and the animal across the two questions (Where is the purple rabbit? → Now, where is the pink monkey?). While the first questions had no particular prosodic emphasis, the second questions were presented either with or without a pitch prominence on the color (e.g., ORANGE/orange lion). Half of the fillers repeated the color with or without a pitch prominence on the animal (orange rabbit→orange FISH/fish), whereas the other half switched both the color and the animal with or without the prominence on the animal (purple turtle→green SQUIRREL/squirrel).

Nine individuals with WS and 9 CA control participants (age: 10-35) showed striking similarities in their fixation patterns despite clear differences in verbal mental age (WS: 4;04-13;00 vs. CA: 10;04-18;06) and IQ (WS: 40-95 vs. CA: 99-126). In the sequences that repeated the animal, the pitch prominence on the color (pink lion→ORANGE lion) speeded the looks to the target cell (lion) in both groups ($p<.01$). Crucially, WS showed no delay in the facilitative effect of prosody (confirmed in 0-300ms window from the noun onset: Fig.1&2 left); both groups looked at the target immediately after they heard the prominence on the color. When the color and the animal switched, prominence on the color (purple rabbit → PINK monkey) led to incorrect fixations to the initially mentioned animal (rabbit), which delayed the fixations to the correct target (monkey): the split between the red lines was later than the split between the blue lines (Fig.1&2: right), although this delay was statistically significant only in CA (300-600ms). In sum, the data show that WS individuals are sensitive to prosodic cues: they interpret pitch prominence as a contrastive cue and make immediate anticipatory eye-movements. The results necessitates a reevaluation of past prosody-assessment tasks that suggested 'clearly lower' prosodic performances in WS.



Example Slide

Fig1&2: line colors
Red: pitch prominence
Blue: no prominence



Animal repeated Animal & Color Switched

Fig.1 CA fixation prop.
Filled = target
Open =initial animal

Fig.2 WS fixation prop.
Filled = target
Open =initial animal

The processing of prosodic focus in French

Jui Namjoshi (University of Illinois at Urbana-Champaign) & Annie Tremblay (University of Kansas)
namjosh2@illinois.edu

French listeners use prosody to segment speech into words [1] and break down sentences into syntactic constituents [2]. However, relatively little is known about French listeners' use nuclear vs. contrastive accents in speech comprehension [3]. Although focus in French is typically conveyed by a combination of syntax (e.g., clefting) and prosody (contrastive accent), it can also be signaled by only prosody, with contrastive pitch accents occurring on the first syllable of focused constituents (cf. nuclear pitch accents, which occur on the last non-reduced syllable of the Accentual Phrase) [3,4]. Native English listeners use contrastive accents to anticipate upcoming referents in the sentence [5]. Given the common use of syntax to express focus in French, it is unclear whether French listeners also use contrastive accents alone to anticipate upcoming referents.

20 French listeners completed a visual-world eye-tracking experiment. Stimuli included a context sentence with an object and a critical sentence with another object. Critical sentences were manipulated in a 2x2 design, with accent type on the object (nuclear/contrastive) and information status for the person owning the object (new/given) as within-subject variables (see (1)-(2)). Results were analyzed in the ambiguous time window (i.e., from object noun onset to person noun onset) and in the post-disambiguation time window (i.e., from person onset to person offset). Mixed-effects models on the proportions of target and competitor fixations were run in both time windows, with accent type, information status for the person, and their interaction as fixed variables, and with participant and item as random variables. Results showed a significant accent-by-information-status interaction on the proportions of competitor fixations in the post-disambiguation time window (p<.05). Further tests revealed that contrastive accents yielded lower proportions of competitor fixations when the person was given than when the person was new. This suggests that contrastive accents on the object in the critical sentence lead French listeners to anticipate a given person referent in the sentence.

These findings indicate that like in English, contrastive accents constrain sentence interpretation in French, even when focus is not expressed syntactically. Hence, prosody may play a more important role in French listeners' processing of focus than previously thought.

(1) (Context) sentence 1:
Clique sur le macaRON de Marie-Hélène. 'Click on the macaron of Marie-Hélène.'

(2) (Critical) sentence 2:
a. *Clique sur le chocoLAT de Marie-Hélène.* (nuclear pitch accent, given person)
b. *Clique sur le CHOcolat de Marie-Hélène.* (contrastive pitch accent, given person)
c. *Clique sur le chocoLAT de Jean-Sébastien.* (nuclear pitch accent, new person)
d. *Clique sur le CHOcolat de Jean-Sébastien.* (contrastive pitch accent, new person)
'Click on the chocolate of Marie-Hélène/Jean-Sébastien.'

References:

[1] Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language*, *51*(4), 523-547.
[2] Grosjean, F. (1996). Using prosody to predict the end of sentences in English and French: Normal and brain-damaged subjects. *Language and cognitive processes*, *11*(1-2), 107-134.
[3] Féry, C. (2001). Focus and phrasing in French. *Audiatur vox sapientiae: A Festschrift for Arnim von Stechow*, 153-181.
[4] Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (Ed.), Intonation: Analysis, modeling and technology (pp. 209-242). Boston: Kluwer.
[5] Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, 58(2), 541-573.

Scalar implicature: a whirlwind tour with stops in processing, development and disorder

Jesse Snedeker (Harvard University)
snedeker@wjh.harvard.edu

Over the past ten years, there has been avalanche of work on scalar implicature in adults and children. I suspect this popularity is driven by two factors. First, the phenomenon is pragmatic, yet crisp enough to seem tractable to hard-nosed experimentalists. Second, the initial psycholinguistic studies (Bott & Noveck, 2004; Huang & Snedeker, 2009a) revealed failures which looked a bit like those that had supported modular theories of parsing during the 1980's and 1990's, creating a rhetorical opening for constraint-based theories.

Here, from my perspective, is what we have learned.

On processing: Psycholinguistics is built on the hypothesis that computation in biological systems unfolds over time--a hypothesis that we lean on and validate whenever we use time as a dependent measure. Scalar implicatures, on all theories, involve computations that go beyond semantic meaning. Thus implicatures should require time to compute, and in many studies they do (e.g., Hartshorne & Snedeker, 2009; Bott, Bailey & Grodner, 2012). What about those cases where they appear to be instantaneous (e.g., Grodner et al., 2010; Breheny et al., 2012)? Our standard theoretical kit contains two tools that might explain this: 1) the computations could be stored and retrieved (lexicalized); 2) the context could provide enough information to allow the computation to be done ahead of time. In the case of scalar implicature, the data and the nature of the problem favor this second possibility.

On development: Most the developmental studies use tasks in which children are asked whether an over informative statement is acceptable ("Some elephants are animals"). Children, unlike adults, often accept these statements, suggesting that they do not calculate implicatures (Noveck, 2000). However, by 5 years of age, they also recognize that these statements are not as praiseworthy as informative statements (Katsos & Bishop, 2010). One possibility is that children calculate implicatures but are just more tolerant of pragmatic violations. Pragmatic tolerance, however, fails to account for children's failures in tasks that do not require judgments (e.g., Huang & Snedeker, 2009b; Huang, Spelke & Snedeker, 2013). While the jury is still out on what drives development, the correct theory will have to be one: that accounts for the differences across tasks and lexical items, that is consistent with a prolonged period of gradual change, and that isolates the component processes involved in making an implicature and identifies which are sufficient for each task.

On disorder: Children with SLI have pragmatic abilities that are either unimpaired or at least more preserved than their grammatical abilities. In contrast, persons with autism have clear pragmatic deficits but often have syntactic and lexical abilities comparable to typical developing folks. But, curiously, children with SLI show impairments in calculating implicatures (Katsos, Andrés Roqueta, Clemente Estevan & Cummins, 2010), while persons with autism do not (adults: Pijnacker, Hagoort, Buitelaar, Teunisse & Geurts, 2009; adolescents: Chevallier, Wilson, Happé & Noveck, 2010; children: Hahn, Huang & Snedeker, in prep). This provides support for any theory which draws a distinction between two classes of pragmatic processes, those which are closer to semantics and those which draw more broadly on world knowledge and social cognition (Levinson, 2001; Chierchia, 2006; Carston & Hall, 2011). It also suggests that the developmental changes in implicature calculation may be driven by improvements in language processing.

Thursday, March 13
Poster Session I Abstracts

Abstract structure is active during comprehension of collocations

Erin Conwell (North Dakota State University) & Jesse Snedeker (Harvard University)
erin.r.conwell@ndsu.edu

In some paradigms, structural priming during comprehension extends across sentences with different lexical content, indicating that abstract syntactic or semantic structures are involved in comprehension (e.g., Thothathiri & Snedeker, 2008). However, there is also evidence for verb-specific structural priming (Arai, Van Gompel, & Scheepers, 2007), suggesting that syntactic alternations are stored with or accessed via the verb. The present study explores another way in which structure might be lexicalized. Dative sentences tend to contain pronominal arguments with high frequency verbs. The double-object dative (DO) often contains a pronominal recipient, making "give me" a common collocation in that structure, while the prepositional dative (PP) occurs with pronominal themes, making "give it" frequent. High token frequency for a few types is associated with storage and low productivity (e.g., Baayen, 1992). Indeed, a Bayesian model of language learning indicates that a rational learner would store a "give me X" version of the DO-construction, and a "give it to Y" version of the PP-construction (Conwell, O'Donnell & Snedeker, 2010).

We used the visual-world act-out task from Thothathiri and Snedeker to explore whether priming from high-frequency collocations (likely to be stored) is different than priming from low-frequency verb-noun combinations (unlikely to be stored). The target sentences always contained low-frequency verb-noun combinations. If high-frequency dative sequences are stored, and if priming depends on taking the same path to the structure, then priming should be absent or reduced when the prime uses a frequent verb-pronoun collocation rather than an infrequent verb-noun combination. In contrast, if such forms are not stored, or if priming depends on the structure itself rather than the pathway of access, priming should be equivalent.

Participants were 96 adult native speakers of American English. In the DO-NP and PP-NP conditions, all prime sentences contained only full noun phrases. In the DO-Pronoun condition, all prime sentences had *me* as the recipient (e.g., *The lion says, "Give me the cheese."*) and in the PP-pronoun condition, all prime sentences had *it* as the theme (e.g., *The fish says, "Give it to the lion."*). The pronoun and NP conditions were otherwise identical and, in both cases, the referent of the first argument was always mentioned in the previous sentence.

Each prime trial was followed by a target trial with a different display. All participants heard the same targets with full NP arguments: four PP's and four DO's. There was a temporary ambiguity in the identity of the first post-verbal argument with one cohort member being a potential recipient (lion) and the other a potential theme (light). We predicted that participants who heard PP primes would look more to the object than to the animal, while participants who heard the double object dative would show the opposite pattern.

We calculated the difference in looking time to the critical object and critical animal in the window 200-600ms after the onset of the first post-verbal argument. There was a main effect of prime dative (PP vs. DO, $p<.001$), but no effect of argument type (*pronoun vs. NP,* $p=.74$). We also found a significant interaction of prime dative and argument type ($p=.046$). Surprisingly, primes that contained a pronoun produced stronger priming effects than those with full NPs, inconsistent with the expectation that high frequency collocations would be weaker primes.

Our findings indicate that, during language comprehension, structural priming is not reduced for collocations; predictable sequences activate the broader PP and DO constructions at least as effectively as less predictable ones. This could suggest that "give me" and "give it" are not stored, but it could also indicate that priming during comprehension is a product of the structure itself, rather than the pathway to the structure. After all, in language production, abstract structural priming is present even for idioms (Konopka & Bock, 2009). The slight boost in priming from pronominal forms could indicate that ease of processing or typicality of the referential form facilitates priming (Bresnan & Nikitina, 2009).

Encoding and retrieval interference in dependency resolution

Jens Roeser, Lena Jäger, Lena Benz, & Shravan Vasishth (University of Potsdam)
jroeser@uni-potsdam.de

Introduction: Structurally inaccessible noun phrases (NPs) which are not in a binding relation with the anaphoric element have sometimes been found to be incorrectly bound by the anaphor, leading to a slowdown in processing (Badecker and Straub, 2002; Patil et al., 2014). This so-called inhibitory interference effect occurs when the legal antecedent and the structurally inaccessible NP share features (e.g., gender in English reflexives *himself*/*herself*). Dillon (2011) attempted to explain this slowdown not in terms inhibition but rather in terms of encoding interference: in configurations where the syntactically licensed antecedent and the structurally inaccessible NP share a feature, a degradation might occur of the memory trace of the legal antecedent due to mechanisms like feature overwriting (Oberauer and Kliegl, 2006), causing increased processing time during retrieval of the legal antecedent. In this experiment, we aim to determine whether interference effects in anaphoric dependency processing can be explained by encoding interference.

Swedish possessive pronouns (*hans* 'his') show gender agreement with the antecedent, while possessive reflexives (*sina* 'his'/'her') do not (see Table 1). If an inaccessible NP matches in gender with the legal antecedent, in the case of gender marked pronouns either encoding or inhibitory interference could explain any slowdown seen at the pronoun; by contrast, in possessive reflexives, any slowdown seen can only be attributed to encoding interference, because the reflexive does not have any gender marking which therefore cannot be a retrieval cue. Thus, if only inhibitory interference explains any slowdowns at the anaphor, an interaction of anaphor type and interference is expected: a slowdown in the gender marked pronoun due to feature match, but no slowdown in the possessive reflexive. If only encoding interference accounts for the slowdown, a main effect of interference is predicted, with no interaction of interference with anaphor type.

Method: 32 Swedish natives read sentences manipulated for anaphor type and interference (Table 1). Eye movements and response accuracy requiring anaphor resolution have been measured.

condition	antecedent	region 2	region 3	region 4	pre-critical	critical	spill-over	wrap-up
pron (match/ mism.)	Åke	säger	att	**Alf**/ **Eva**	jobbade med	**hans**	sysslingar	på helgerna.
	Åke [M]	says	that	Alf [M]/ Eva [F]	worked with	his [M]	siblings	at the weekend
refl (match/ mism.)	Åke	som	**Alf**/ **Ann**	tackade	ringer	**sina**	sysslingar	på kvällen.
	Åke [M]	who	Alf [M]/ Ann [F]	thanked	calls	his [∅]	siblings	in the evening

Table 1: Stimulus sentences with regions of interest.

Results: Question-response accuracy revealed an interaction of interference and anaphor type (p<0.01) showing a lower response accuracy due to interference in pronouns (p<0.0001) but not in reflexives (p=0.71). Re-reading time regressive, the sum of all second-pass fixation durations in the pre-critical region after a region to its right has been fixated, shows an interaction of interference and anaphor type (t=2.18) with a slowdown in pronouns (t=2.16), but not in reflexives (t=-0.94). In the pre-critical region, last pass reading time (the sum of all fixation durations during the last pass) shows an interaction of interference and anaphor type (t=2.11) with a slowdown in pronouns (t=3.01) but not in reflexives (t=0.03).

Conclusion: Interference effects seem to be due to inhibitory processes following incorrect initial retrieval of the structurally inaccessible NP and not due to encoding interference.

References

• Badecker, W. and Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28:B13–B26. • Dillon, B. W. (2011). *Structured access in sentence comprehension*. PhD thesis, Maryland. • Oberauer, K. and Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55:601–626. • Patil, U., Vasishth, S., and Lewis, R. L. (2014). Retrieval interference in syntactic processing: The case of reflexive binding in english. MS under revision.

Exploring socioeconomic differences in syntactic development through processing

Yi Ting Huang, Kathryn Leech, & Meredith Rowe (University of Maryland College Park)
ythuang1@umd.edu

Prior research has found that a child's socioeconomic status (SES) background is a strong predictor of language development. Quantitative differences in the parental input of low- and high-SES families lead to striking gaps in early vocabulary development [Hart & Risley, 1995]. Nevertheless, while the effects of SES on vocabulary have been well-studied, those on syntax are less understood. The current study recruits the active/passive alternation – an important test case for exploring syntactic development – to examine possible SES differences. Previous research has found that children have more difficulties with passives compared to actives. Two major hypotheses have been proposed to explain why. One possibility is that asymmetries in proficiency reflect the paucity of passives in children's input. Another possibility is that they reflect properties of passives that are problematic for processing. Sentence (1) illustrates that passives cannot be distinguished from actives until after the verb. This creates a temporary ambiguity in the grammatical role of NP_1. Also since passives are less frequent than actives, children may initially misanalyze NP_1 as an agent but later fail to revise their interpretation after the verb. These accounts differ in their predictions of the scope of difficulties that are experienced in a population that hears considerably less parental input for passives.

We investigated the effects of SES differences on the comprehension of passives. Five-year-olds (n=69) from low- and high-SES families heard sentences like (1) and (2) while their eye-movements and actions were measured to displays featuring expressed items (SEAL), likely agents (SHARK), and themes (FISH). Unlike (1), the referent of NP_1 in (2) is initially ambiguous. Previous work has found that this ambiguity prevents children from automatically interpreting NP_1 as an agent, allowing role assignment to be postponed until after the disambiguating verb [Huang et al., 2013]. If the processing demands create difficulties in interpretation, then children from low-SES families should perform worse than their high-SES counterparts for passives that require revision like (1). However, they should perform similarly for passives that do not require revision like (2). In contrast, if difficulties with passives reflect the frequency statistics of this construction, then children from low-SES families should perform worse with both (1) and (2).

Analysis of eye-movements focused on predictions of pronoun identity as measured by looks to likely agents minus likely themes. All children used verb morphology to incrementally assign grammatical roles. In Expressed NP_1 trials, they looked to likely themes for actives and likely agents for passives. In Pronoun NP_1 trials, this pattern appropriately reversed. Analysis of actions focused on responses depicting correct role assignments between expressed items and inferred referents. These data revealed SES effects on interpretations. All children generated correct actions in both active trials and in the Pronoun NP_1 passive trials. However, they had more difficulty doing so in the Expressed NP_1 passive trials, particularly those from low-SES families. This led to an interaction between NP_1 status, construction, and SES (p<.05). Altogether, these findings demonstrate the presence of SES effects on syntactic development and suggest that differences may reflect an interaction between the processing demands of a construction and the language experience of a child. For passives, all children make statistical predictions of role assignments, but input may facilitate the revision of initial misinterpretations.

(1) Expressed NP_1 trials: Active / Passive The seal is quickly <u>eating it</u> / <u>eaten by it</u>
(2) Pronoun NP_1 trials: Active / Passive It is quickly <u>eating the seal</u> / <u>eaten by the seal</u>

SES background	Expressed NP_1 trials		Pronoun NP_1 trials	
	Active	Passive	Active	Passive
Children from low-SES families	68%	26%	59%	61%
Children from high-SES families	73%	47%	68%	59%

Homophone disambiguation & the social identity of the speaker

Julie Boland & Maryam Seifeldin (University of Michigan)
jeboland@umich.edu

Do social properties of a speaker (e.g., age, gender) influence how easily the alternative meanings of a homophone become accessible? The answer might be "yes" for homophone meanings with strong social biases. For example, people associate the fingernail meaning of "nails" more strongly with women and the hammer meaning more strongly with men. We tested the hypothesis that listeners would fixate an image corresponding to a socially-biased homophone meaning more rapidly when the homophone was spoken in a voice congruent with the social bias. There are at least two reasons why such a pattern might be found. First, social properties of the voice might constrain homophone meaning similar to biasing sentential context, with top-down expectations modulating the accessibility of alternative meanings. Alternatively, acoustic properties correlated with social identity might be lexically encoded via salient exemplars of the words, providing a bottom-up mechanism for a speaker congruency effect. Under the exemplar account, a woman's utterance of "nails" might activate "nails" exemplars corresponding to the fingernails meaning more readily than a man's utterance of "nails" due to acoustic similarity between the current input and stored exemplars. We designed a sequence of experiments to measure the impact of each of these mechanisms on spoken homophone recognition.

Three visual world experiments measured eye movements to the referent of a spoken homophone when the voice was either consistent or inconsistent with the pictured meaning of the homophone. We selected 24 homophone meanings that had a strong bias on either the male/female dimension or the child/adult dimension. For example, in Experiments 2 and 3, we measured the time to fixate a fingernails image when "Look at the nails" was spoken in either a woman's voice or man's voice. Only one homophone meaning was depicted on any given trial; it appeared with three other images (one phonologically related, two unrelated) at the onset of the spoken homophone in all three experiments. Across experiments, we manipulated the carrier phrase in order to evaluate the proposed top-down and bottom-up mechanisms for a speaker congruity effect.

Design & Predictions. In Experiment 1, the carrier phrase was a biasing sentence context (e.g. "At the end of the month, I write out a check" in either an adult voice or a child voice). The bias of the sentence was always consistent with the depicted homophone meaning. Experiments 2 and 3 used the neutral carrier phrase "Look at the…" with or without an overlay of crowd noise. The bottom-up mechanism predicts consistent speaker congruity effects across experiments, because socially congruent voices produce tokens that more closely match stored lexical exemplars. In contrast, the top-down mechanism should result in larger speaker congruity effects when listeners are encouraged to develop expectations about upcoming referents. In our study, only Experiment 1 fostered listener expectations, with homophone meanings somewhat predictable from the linguistic context. In Experiments 2 and 3, the carrier phrase contained no clues to the upcoming target word, beyond the very general social cues carried by the voices. In all three experiments, half of the trials were consistent with the social cues and half were inconsistent.

Results. Experiment 1 employed a passive looking task, while Experiments 2 and 3 employed directed action tasks and had higher overall rates of target fixations, as well as longer-lasting target fixations. Consistent with the proposed top-down mechanism, we found speaker congruity effects only in Experiment 1. We found no evidence of exemplar-based lexical activation.

How sentence processing benefits from the inflectional richness of a language

Sophia Manika (UiL OTS, Utrecht University)
s.manika@uu.nl

Consider: (1) John praises/helps, very often, Maria and not his other friends.

Despite the similarities between *praise* and *help*, in any relevant aspects, it has been found that processing '*helps'* is faster than processing '*praises'* both *in isolation* (Milin et al., 2009 a.o.) and *within* a sentence (Manika et al., 2013). This contrast is predicted by the value of the verbal paradigm's *inflectional entropy* (H). H, the sum of the frequency/function ratio of each inflected form, is an information-theoretic measure that describes the distribution of the inflectional paradigm and has been shown to correlate with an inflected form's processing speed. High H (HH, e.g. *help*) describes more 'uniform' distributions that support the activation of a verb form, facilitate its retrieval and require fewer resources. Low H (LH, e.g. *praise*) describes less uniform distributions that do *not* support the form's activation and delay the form's retrieval, requiring, thus, more resources. Consequently, more availability in resources makes *Maria* to be processed faster as an object of '*helps'* (HH verb) than as an object of '*praises'* (LH verb) (Manika et al., 2013). Interestingly, morphologically rich*er* languages (e.g. Greek) have verbs with higher values of H than morphologically poor*er* languages (e.g. Dutch) that range in lower values of H. Since an association between inflectional entropy and the amount of resources consumed during sentence processing is assumed, in a sentence like (1), all other factors kept constant, the verb *praise* would be read faster in Greek ('*epeno'*) than in Dutch ('*prijzen'*) because Greek is a HH language and Dutch is a LH language.

The present paper investigates the connection between inflectional entropy and processing effort/resources and it crucially demonstrates that sentence processing speed benefits from rich morphology. More precisely, Greek verbs that have overall higher values of H, because of Greek's rich inflection, were retrieved faster than Dutch verbs that have lower values of H, because of the poor inflection. This 'saving' on processing resources in the case of Greek (HH) resulted in *Μαρία* being processed faster than *Maria*.

Forty-two native Dutch and forty-one native Greek students read sentences like (1) in self-paced reading experiments. Words between the verb and the PN were kept constant across items. The PNs in subject and object position were controlled for frequency and number of syllables. The inflectional entropy of the matrix verb was varied and the values between the two languages did not overlap (the highest value of H of a Dutch verb was lower than the lowest value of H of a Greek verb). Reading times (RT) in each word were analyzed with LMER analyses using a 2-stage model (Fine et al., 2013). In the first stage log-transformed RTs were regressed against the word's length and position in the sentence and the item's position in the task. The residuals of that model served as the dependent variable for the second stage and were crossed with language (Dutch vs Greek) and object type (PN vs reflexive, which served as a baseline because its processing is known to be fast). Results are presented per region. *Verb region*: HH verbs were retrieved faster than LH verbs ($t = -2.758$, $p<.05$). *Object integration*: The main effect of object type on RTs was significant at the region of the object ($t = -2.458$, $p<.05$). Sentences with PNs were read significantly slower than those with reflexives. There was a significant interaction between object type and language ($t = 2.476$, $p<.05$). Simple slopes reveal that HH language (Greek) accelerates RTs for sentences with PN ($t = -3.342$, $p<.01$).

In a nutshell, **i**nflectional entropy is tightly connected to the morphological richness of a language that in turn affects the speed of sentence processing. Processing of a PN requires a costly introduction of a new entity in the discourse. In a morphologically poor language, like Dutch, in which verb retrieval already consumes a lot of resources, introducing a PN is even costlier. The fact that the reflexive condition was processed equally fast in both languages provides us with a baseline. Hence, I can conclude that the delay of Dutch PN *Maria* is a result of the shortage in available resources as predicted by the fact that Dutch, being a morphologically poor language, has lower values of inflectional entropy.

Experience with dialectal variants modulates online syntactic comprehension

Scott H. Fraundorf & T. Florian Jaeger (University of Rochester)
scottfraundorf@gmail.com

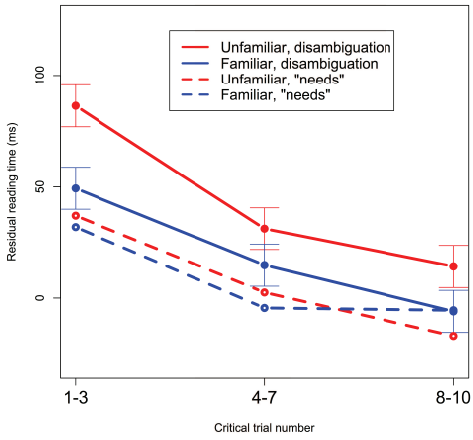
What are the consequences for online syntactic processing of long-term experience with dialects differing in their syntactic constructions? Infrequent structures are harder to process [1] and repeated exposure to structures in the lab decreases processing difficulty [2,3], suggesting that individual experience affects processing. It is as of yet unknown, however, how long-term experience with dialectal syntactic variants affects online sentence processing.

We compared comprehension of the *needs* + participle construction in 'Pittsburgh' English [4], which uses (1) below to mean (2), by readers whose dialects did or did not include this construction. For readers unfamiliar with the construction, the string *The display case needs polished…* can only be interpreted if *polished* is expected to modify an upcoming noun, such as *glass* in (3). If this expectation is violated by the participle continuation in (1), processing should be slowed, providing a measure of the degree to which *needs* + participle was unexpected [5].

(1) The display case needs polished before we add the trophy. **[needs + participle]**
(2) The display case needs to be polished before we add the trophy. **[conventional]**
(3) The display case needs polished glass to look as good as possible. **[modifier]**

We recruited participants (*N* = 173) for a web-based experiment using Amazon Mechanical Turk to target regions in which participants were likely (Ohio and western Pennsylvania) or unlikely (Colorado) to have heard the construction. A post-experiment survey explicitly assessed dialectal background; filler questions about other, unpresented structures controlled for any tendency to affirm all dialectal variants. We identified 129 *familiar* participants with prior knowledge of the construction and 44 *unfamiliar* participants without.

All participants self-paced word-by-word moving window reading of short 20 paragraphs, each 3-4 sentences, constructed to simulate e-mail messages (as per the cover story of the experiment). Ten critical paragraphs (15% of total sentences) contained the *needs* + participle construction in the next-to-last sentence, as in (1).



The figure displays length- and screen position-corrected residual RTs for the disambiguating region, such as *before we* in (1), and for the unambiguous word *needs* as a function of participant group and trial number. We analyzed reading times using a linear mixed model with the maximal random effects structure justified by the data. A group x region interaction (t = 4.50, $p < .001$) indicated that participants *familiar* with *needs* + participle read the disambiguation faster than participants *unfamiliar* with it. This group difference did not occur at the unambiguous *needs* (t = 0.47, p = .64), indicating it was not due to differences in overall reading speed. A group x region x trial interaction (t = 2.06, $p < .05$) indicated that the effect at *needs* diminished after the first three trials as the unfamiliar participants read the *needs* + participle construction more like the familiar participants.

These results indicate that online syntactic comprehension is sensitive to differential experience with dialectical variation, but that readers quickly learn to better comprehend unfamiliar constructions (as in [5]). Both results suggest a strong role of experience in syntactic processing. The present work also demonstrates that web-based methodologies can target recruitment of participants varying in dialectal features relevant to online sentence processing.

References

[1] MacDonald, Pearlmutter, & Seidenberg. 1994. *Psychological Review*. [2] Fine, Jaeger, Farmer, & Qian. 2013. *PLoS ONE*. [3] Wells, Christianson, Race, Acheson, & MacDonald. 1999. *Cognitive Psychology*. [4] Murray, Frazer, & Simon. 1996. *American Speech*. [5] Kaschak & Glenberg. 2004. *Journal of Experimental Psychology: General*.

Language processing shapes language change: Redundant cues are less likely to be maintained in language

Maryia Fedzechkina (University of Rochester), Elissa L. Newport (Georgetown University),
T. Florian Jaeger (University of Rochester)
mashaf@bcs.rochester.edu

Languages around the world share a number of commonalities known as *language universals*. Our research examines the hypothesis that the cross-linguistic distribution of grammars can at least in part be accounted for in terms of their processing or communicative utility (e.g., Bates and MacWhinney, 1982; Hawkins, 2004). We ask in particular whether the well-documented correlation between word order (WO) freedom and the presence of a case system in a language (languages with flexible WO typically have a case system, while languages with fixed WO typically do not) can be explained by a bias to balance production effort and informativity of cues to grammatical function.

Present work addresses this question using a miniature artificial language learning paradigm in which we present adult learners with experimentally created miniature languages that use several competing forms to express the same meaning and study learners' deviations from the input. Two groups of participants (all monolingual native speakers of English) were exposed to one of two miniature artificial languages in three 35-minute sessions spread over three days. Both languages had optional case-marking (67% of objects were case-marked; subjects were never case-marked). Language 1 had fixed WO (100% SOV). Language 2 had flexible WO (50% SOV; 50% OSV). Since grammatical function assignment could be unambiguously identified by WO in Language 1, case was not necessary for successful recovery of sentence meaning but added redundant and clear marking. In Language 2, WO was uninformative about grammatical function assignment, and case-marking, when present, provided unique information about sentence meaning.

We hypothesized that if learners indeed trade off the information content of a cue and the amount of effort necessary to produce this cue, they should restructure the input and use less case-marking when WO is highly informative of grammatical function assignment (fixed WO language). The results support this hypothesis (Figure 1): learners maintained case-marking only in the flexible WO language (71% of all objects were case-marked in learners' productions on the final day of training), but not in the fixed WO language (41% of all objects were case-marked in learners' productions on the final day of training).

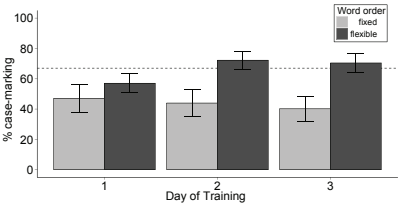


Figure 1

Importantly, learners of the flexible WO language did not regularize case-marking in the language overall, but instead conditioned case on WO. Even though the two orders were case-marked equally often in the input, learners produced significantly more case-markers in OSV sentences (Figure 2), which provides an earlier disambiguation point (cf. Maximize Online Processing, Hawkins, 2004). These results suggest that preferences associated with incremental sentence processing influence the development of languages over time.

These learning outcomes parallel natural typological phenomena: the presence of case systems in languages with flexible WO and their absence in languages with fixed WO and patterns of diachronic change such as that from the Old English (a language with flexible WO and rich case-marking) to Modern English (a language with fixed WO and only rudimentary case-marking).

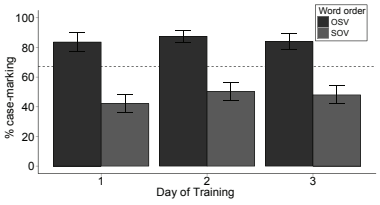


Figure 2

Listeners maintain and rationally update uncertainty about prior words in spoken comprehension

Klinton Bicknell (University of Rochester, Northwestern University), Michael K. Tanenhaus (University of Rochester), & Florian Jaeger (University of Rochester)
kbicknell@northwestern.edu

Recent work suggests that sentence processing may be characterized as rational inference over noisy probability distributions (Levy et al., 2009; Dahan, 2010; Gibson et al., 2013). But to what extent do listeners maintain and update uncertainty about previous words as an utterance unfolds, and are these updates rational? While there is some evidence that listeners maintain distributions over word candidates for several syllables (Connine et al., 1991; Dahan, 2008; Johnstone et al., 2013), there may be severe limits on this process inconsistent with rational inference. Here, we build on Connine et al. and show that uncertainty is still maintained 6 syllables later. Moreover, our results suggest that uncertainty is not restricted to maximally ambiguous tokens, but that acoustic and contextual information combine rationally.

Connine et al. created a 6-step VOT continuum between the words [dent] and [tent], with unambiguous endpoints and 4 ambiguous steps near the category boundary. Listeners identified [dent]/[tent] in sentence contexts, where the left context provided no information about word identity and the right context disambiguated between dent/tent. *VOT* (6 steps from [dent]-[tent]) was crossed with disambiguating *context* (tent vs. dent) and *distance* between [dent]/[tent] and the disambiguating right context (1, 3, 6 syllables; see example). They found a main effect of context, suggesting that comprehenders maintain uncertainty about prior words and update their beliefs using following words. However, Connine et al. also reported two limits inconsistent with fully rational updating: (1) no effects were found for the 6-syllable distance and (2) an interaction of VOT and context in an ANOVA over proportions suggested that uncertainty is only maintained for the most ambiguous VOTs. Given that relevant context is often delayed and category-boundary VOTs are rare, these conclusions limit the usefulness of belief updating.

We raise two issues with these conclusions. Connine et al. allowed comprehenders to respond at any time: 85% of responses in the 6-syllable condition were made *prior* to the disambiguating context. Thus, the absence of 6-syllable condition effects (1) may reflect limited power. Regarding (2), we note that an optimal integrator combines acoustic and contextual information additively in log-odds, which yields an interaction in an ANOVA over proportions.

We conducted two web-based experiments (N=80) modeled after Connine et al. Crucially, participants could only respond after the full sentence. We replicated the finding that comprehenders maintain and update uncertainty about prior words. We also found a reliable effect of context in the 6-syllable condition, demonstrating that uncertainty is maintained for at least 6 syllables. We also replicate the interaction of VOT and context in an ANOVA over proportions. Crucially, however, this interaction disappears when using more appropriate analyses in log-odds space (ANOVA on empirical logits, logistic mixed models). Additional analyses allowing non-linear VOT effects also revealed no evidence that context effects are larger for maximally ambiguous stimuli. These analyses argue that acoustic and contextual information are additive in log-odds space, as predicted by a rational integration model.

In conclusion, we find that listeners maintain and update uncertain beliefs about previous words. Contrary to previous findings, we show that this process bears the signatures of rational integration: (1) Uncertainty is maintained and updated even 6 syllables downstream and (2) acoustic and contextual cues combine additively in log-odds space. These results support models of sentence processing as rational inference over noisy perceptual information. Moreover, processes of maintaining and updating uncertainty about previous words required for rational inference are less compromised by cognitive limitations than previously believed.

Example: (with both dent/tent disambiguating context in italics)
1 syllable: When the __ *corroded/collapsed* … *3 syllables:* When the __ in the *fender/forest* …
6 syllables: When the __ was noticed in the *fender/forest* …

Learning language from the environment depends on the fitness of both the learner and the environment

Shiri Lev-Ari (Max Planck Inst. for Psycholinguistics) & Sharon Peperkamp (ENS,EHESS,CNRS)
shiri.lev-ari@mpi.nl

Speakers learn language from their environment. This learning continues throughout speakers' lives as they continuously monitor the environment and adjust their representations accordingly. In general, learning from the environment is better if (a) learners make over-all adjustments to permanent and representative changes in the environment but not to transient changes, and if (b) learners adjust more the poorer their fitness in the environment is. We test whether the learning of linguistic input follows these general tenets.

Study 1 tests hypothesis (a) with a perceptual learning paradigm. Native French speakers listened to native (representative) or Dutch accented (unrepresentative) French with imperceptibly manipulated VOTs. Results show that listeners adapt to the VOTs of both native and non-native speakers, but only generalize their learning to a new native speaker if they learned from the native speaker. This shows sensitivity in learning to the representativeness of the learning environment and its relevance to other linguistic contexts.

Studies 2-3 test hypothesis (b) by examining whether less fit speakers, as approximated by poorer performance on a task, learn the linguistic patterns in their environment more. Study 2 tests this by analyzing the grammatical structure used by players in the TV game show Jeopardy. Players could use one of two grammatical structures when requesting a clue - with and without a preposition (*Natural Wonders for/∅ 200*). Analyses show that the more money players had won, the less likely they were to repeat the structure used by the previous player. Analyses further showed that the preference to repeat the frequent (with a preposition) over the infrequent (∅) structure is reduced when winning more money, suggesting reduced reliance on linguistic frequency with better fitness. Study 3 provided further support from a game task in the lab, where small groups of French speakers played a modified version of the game Go Fish. The modified version required participants to describe the cards they were requesting. As with Jeopardy, one property of the cards, the flavor of the depicted ice-cream, could be described in one of two grammatical structures - with and without a proposition (*la glace au/∅ chocolat*). Analysis of participants' utterances revealed that participants were more likely to repeat the previous player's grammatical choice the more cards their addressee had taken from them.

Studies 2-3 then show that speakers' adaptation to the grammatical structures used in the environment depends on how well they perform in the environment. This dependence occurs even though the linguistic choices are not directly linked to fitness in the environment, as success in neither Jeopardy nor Go Fish depends on grammatical choices.

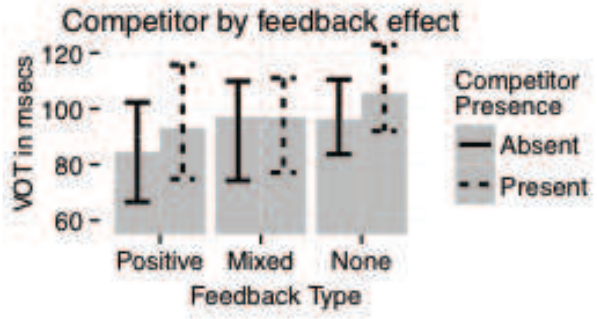
Together, this set of studies shows that the degree to which individuals adjust their language according to encountered input varies and depends on their own as well as the environment's fitness. This indicates that the weight given to encountered input depends on the circumstances at which it is encountered. These results have consequences for language change as they suggest that innovations are more likely to be adopted during difficult times, when individuals perform more poorly, and by those of a more vulnerable position (studies 2-3) as well as that the learned innovations are more likely to be maintained when these difficult circumstances or poor position seem representative (study 1).

Effects of interlocutor feedback on speaker phonetic production in a simulated-communication task

Esteban Buz & T. Florian Jaeger (University of Rochester)
ebuz@bcs.rochester.edu

Is language production organized for robust communication? Some recent accounts argue that for phonological/phonetic production, the answer is no [1-3]. Under these accounts, evidence that speakers produce contextually more confusable words with a more distinguishable acoustic signal is attributed to production ease [1, 2] or comprehension processes that affect word-specific phonetic representations [4, 5]. This contrasts with accounts in which production is organized to take into account perceived communicative success of previous articulations [6]. Therefore, determining whether speakers' articulations reflect sensitivity to feedback from interlocutors about the success of their previous productions is crucial for distinguishes between these two radically different hypotheses about the architecture of the production system.

In a **web-based simulated-communication task (N=60)** speakers instructed what they believe to be interlocutors to choose one word from among three visually displayed options. Interlocutors were programmed but highly believable (based on post-experiment survey). In critical trials (36 of 90 total), speakers uttered voiceless target (e.g. *pill*) with or without a minimal pair distractor visually co-present (e.g. *bill*, following [4]). A time bar counted down 10 seconds. Speakers were randomly assigned to one of three groups. In the **no feedback** group, trials simply ended, with no indication of what word the interlocutor clicked on. In the **positive feedback** group the interlocutor always chose the correct word (indicated by a green box around the target at the end of the trial). In the **mixed feedback** group 7 trials (thereof, 5 critical) ended with the wrong choice (indicated by a red box around the distractor). Interlocutor response times in all conditions were modeled after previous experiments (incl. item-specific response times and speed-up over the course of the experiment).



VOT for target words was measured for 19 speakers (annotation continues for the remainder) and analyzed using a mixed linear regression (maximal RE structure) with effects for visual co-presence (contrast coded), feedback group (Helmert coded), and their interaction. VOT times were longer for targets uttered with the minimal pair neighbor present ($\beta = 2.99$; $t = 3.85$; $p < .01$), replicating [2]. This effect is attenuated when speakers are given mixed feedback as compared to the other two conditions ($\beta = -1.53$; $t = -2.52$; $p < .05$).

These results replicate existing findings using a web-based paradigm and further find that interlocutor errors affect speaker articulations. The reduction of context differences in the mixed feedback group suggests that speakers might respond to inconsistent feedback with general hyper-articulation. Such a finding supports the claim that speakers are sensitive to interlocutor feedback and modify future articulations based on perceived past communicative success. The paradigm we have developed allows for the collection of large amounts of speech data in less than one day. A complimentary paradigm (currently running) obtains implicit (RT) and explicit (clarity rating) measures of intelligibility for each of these productions, in addition to the commonly used phonetic measures (like VOT, analyzed above).

References Cited

1. Bard, E.G., et al. *JML*, 2000.
2. Arnold, J.E. *LCP*, 2008.
3. Gahl, S., Y. Yao, & K. Johnson. *JML*, 2012.
4. Baese-Berk, M. & M. Goldrick. *LCP*, 2009.
5. Pierrehumbert, J.B., in *Frequency effects and emergent grammar*, 2001.
6. Jaeger, T.F. Front Psychology, 2013.

Error detection in native and non-native speakers provides evidence for a noisy channel model of sentence processing

Robin Hill and Frank Keller (University of Edinburgh)
R.L.Hill@ed.ac.uk

Recent accounts of human sentence processing have given a central role to error-correction mechanisms (e.g., Levy, 2008), formalizing sentence processing using a noisy channel model, in which an intended sentence s_i is transmitted as a (potentially distorted) sentence s_p. The goal of the comprehender is then to recover s_i by combining $P(s_i)$, the prior probability of the intended sentence, and $P(s_p|s_i)$, the likelihood that s_i will be distorted as s_p (Gibson et al., 2013).

Previous work tested the noisy channel model indirectly using reading and judgment tasks (Levy, 2008; Gibson et al., 2013). Here, we investigate its predictions directly by requiring participants to detect and localize errors. Our materials were drawn from the output of a machine translation (MT) system, which naturally contains errors and therefore is ideal for testing the predictions of the noisy channel model. Specifically, we predict that speaker populations (native vs. non-native speakers) will differ in their language model $P(s_i)$, while different types of errors will vary in their distortion probability $P(s_p|s_i)$.

In an eye-tracking experiment, two groups of participants (20 native speakers of English, and 20 highly proficient non-native speakers) were eye-tracked while reading 120 sentences, each of which they saw in one of two versions (error condition or non-error condition). Participants indicated the presence of an error by pressing a response button and localized it by clicking on the erroneous word. The materials included five types of errors typical for MT output, covering lexical, syntactic and semantic violations: Easy Transposition (TE), where the first two letters of a word are switched to produce a non-word, an error that is easy to detect and provides a baseline; Difficult Transposition (TD), where two internal letters were switched to produce an incorrect but legitimate word; Word Order (WO), where the transposition is at the word level with two words being switched; Mistranslation of Tense or Agreement (MT), where there is a violation in verb tense or a mismatch in gender or number agreement; and Mistranslated Lexical Item (ML), where a semantically connected but contextually inappropriate word is substituted.

The results showed significant differences in error detection rates accross error types: TE $>$ WO $=$ MT $>$ TD $>$ ML, which confirms the prediction of the noisy channel model that error types differ in their distortion probability $P(s_p|s_i)$. Native and non-native speakers did not differ significantly in error detection rates or in total sentence reading times. However, an analysis of the reading times on the target word (which either contains an error or not) revealed differences between native an non-native speakers. As the table shows, first fixation times on the target word were sensitive to the presence of an error in native speakers in three out of five error types. In non-native speakers, only the baseline condition (TE) leads to a significant difference. This indicates that non-native speakers have a less reliable language model $P(s_i)$, which slows down error detection. This is confirmed by an analysis of the spill-over region, where only non-native speakers showed significantly more regressions to the target when it contained an error.

Error Type	Native	Non-native
TE	70.1***	82.6**
TD	33.4*	15.5
WO	24.9	9.7
MT	37.4*	8.8
ML	7.9	32.3

First fixation time on target word (error condition minus non-error condition)

References

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*, 8051–8056.

Levy, R. (2008). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (pp. 234–243), Honolulu.

Partial separation of syntactic representations in closely related varieties. Evidence from Swiss German and Standard German

Constanze Vorwerg & Janine Lüthi & Manuel Straessle (University of Bern) & Martin Pickering (University of Edinburgh)
constanze.vorwerg@isw.unibe.ch

Variability is one of the most fundamental and pervasive facets of language, and often speakers have command of different varieties, such as dialects. However, despite the wealth of data explaining linguistic variation – in terms of either social and geographic factors [1, 2] or individual choices and processing [3, 4] – there is a gap between these two traditionally distinct lines of research related to linguistic variation, and the representation of different varieties in memory has scarcely been studied. Moreover, most psycholinguistic research has been conducted with standard language material, and there is a desideratum for studying the individual processing of systematic variability and especially non-standard varieties.

We have addressed the question of whether in bivarietal native speakers, Swiss German dialects and Standard German are represented completely integrated as one language w.r.t. syntactic and lexical processing, or whether there is a partial separation. Research with bilinguals has shown that different languages, even those that are only moderately related, can share syntactic information, as evidenced by priming across languages [5]. Therefore, shared syntactic representations between closely related varieties are to be expected even more. However, the question is whether syntactic priming is as large between varieties of a language as within a variety, or whether varieties may rather be processed similar to different languages. Another question is whether Swiss G. and Standard G. share lexical representations. If so, the same amount of lexical boost in syntactic priming should occur between as within varieties [6].

We conducted two cross-varietal syntactic priming experiments with picture descriptions. Exp. 1 used NP priming (attribute in relative clause vs. adjective) in three variety versions (Bernese/Bernese, Bernese/Standard, Standard/Bernese). Exp. 2 used double-object priming (dative object vs. prepositional object) in four varietal combinations of St.Gallen and Standard G.. Altogether, there were 48 participants in Exp. 1 (16 in each of 3 sub-experiments), and 40 in Exp. 2 (20 in each of 2 sub-experiments). Noun relation between priming and target trials was manipulated in Exp. 1 (identical vs. translation equivalent / semantically related vs. different), and verb relation in Exp. 2 (identical morpheme vs. semantically related).

We submitted the relative frequency data per participant and per item to repeated measures analyses of variance. Results revealed strong variety effects in both experiments, with priming within Swiss G. (Bernese or St.Gallen) being greater than either between varieties or within Standard G.. The lexical boost effect in Exp. 1 was large within Bernese (priming effects: (a) identical nouns: .41, (b) semantically related: .26, (c) different nouns: .10). It was small or non-existent in the other two conditions (priming effects: (a) identical nouns: .10/.04, (b) translation equivalents: .07/.05, (c) different nouns: .04/.03). In Exp. 2, there was only a tendency for a reduction of priming effects in the condition without verb repetition for within Standard German. Again priming was greater within Swiss G. (St.Gallen: .30) than between varieties (.16/.11), and also with Standard G. (.09).

These results suggest that syntactic representations are only partially integrated between varieties, and that Standard German syntactic representations are less accessible than Swiss German ones. Lexical representations do not seem to be shared. Both syntactic and semantic representations might be no more integrated than those of bilingual speakers. Models of language production need to be able to account for selection between varieties and, more generally, linguistic variants.

[1] Trudgill (1974). *Sociolinguistics*. Penguin Books. [2] Wieling et al. (2011). *PLoS ONE*. [3] Bock (1982). *Psychol. Rev.* [4] Cutler et al. (1983). *Nature*. [5] Hartsuiker et al. (2004). *Psychol. Science*. [6] Cai et al. (2011). *JML.*

The interaction of images and text during comprehension of garden-path sentences

Cassie Palmer & Kiel Christianson (University of Illinois at Urbana-Champaign)
Landry1@illinois.edu

The goal of reading is to construct a discourse representation; however, when information in the discourse is confusing or ambiguous, readers are often not able to create an accurate understanding of the text. Research with garden-path (GP) sentences demonstrates that readers may not arrive at the correct, syntactically licensed interpretation despite apparent reanalysis. Good Enough Theory (GET) claims that readers parse sentences incorrectly by developing only "good enough" representations of the structure of ambiguous sentences and thus never return from "down the garden path"(Ferreira et al., 2002; Ferreira & Patson, 2007).

According to Dual Coding Theory (Paivio, 1971, 1990) comprehension and learning are facilitated when people successfully understand information from two sources (visual and verbal) rather than one (cf. Mayer, 2005). Using eyetracking, we tested GET by investigating how reading processes are affected by the presence of nonlinguistic information (imagery) with GP sentences. Both native (L1) and nonnative (L2) speakers of English read GP sentences and viewed imagery that either reinforced the GP (1) or the non-GP (2) interpretation of the sentence. They were then asked to recall whether or not the interpretation of the sentence and the image matched (post-reading, forced-choice task). The manipulated factors were designed in a 2 x 2 x 2 fully factorial experiment: sentence ambiguity (ambiguous vs. unambiguous), imagery (GP interpretation image vs. non-GP interpretation image), and order (image first vs. sentence first). We predicted imagery presented prior to ambiguous language should help guide the readers' attention in the ambiguous text and should help readers avoid being garden-pathed (or reanalyze completely).

LME models revealed that on the post-reading sentence-and-image matching task, native speakers were best at discriminating the meanings of the two sources of information when the sentence was presented first (b=.30, SE=.09, z=3.42, p<.001). The opposite held for nonnative speakers who performed best when the image came first (b=-0.12, SE=0.032, z=-3.87, p<0.001). Both native speakers (b=-2.57, SE=.09, z=-27.73, p<.001) and nonnative speakers (b=-2.19, SE=.07, z=-29.32, p<.001) discriminated image and text sources best when the image reinforced the ultimately correct interpretation. Both native and nonnative speakers performed best when the sentence came first and the image and the sentence reinforced the correct interpretation (native speakers b=-0.46, SE=0.13, z=-3.49, p<0.001; nonnative speakers b=-0.37, SE=0.10, z=-3.49, *p<0.001)*. Only the native speakers were differentially affected by the ambiguity of the sentence in relation to the order of presentation: native speakers performed best when the sentence was unambiguous and the sentence came first (b=1.27, SE=0.17, z=7.09, p<0.001). Although not significant, this pattern was opposite for nonnative speakers: numerically, nonnative speakers performed best when the image came first for both sentence types. We take this as evidence that nonlinguistic imagery is most effective in grounding the meaning of difficult text for less proficient readers; this points to the importance of imagery as a tool for visually grounding the context for ambiguous language.

(1) Sentence: *While Jim bathed the child that was happy and pudgy giggled with delight.*
(2) Sentence: *While Jim bathed, the child that was happy and pudgy giggled with delight.*

Garden-path interpretation image

Non garden-path interpretation image

Bootstrapping into filler-gap: An acquisition story

Marten van Schijndel & Micha Elsner (The Ohio State University)
vanschm@ling.osu.edu

Non-local (filler-gap) syntactic dependencies (e.g. '[The apple] that the horse ate ___ was red.') have long been an object of study for syntacticians due to their apparent processing complexity. This work seeks to computationally model the acquisition of filler-gap comprehension to establish a lower bound on the necessary modelling assumptions. The only similar previous research developed a model of the acquisition of semantic role labelling [1], but its filler-gap performance was never evaluated, and the model does not seem to be consistent with the observed acquisition timeline (seeming to conflate phenomena from different stages of acquisition).

Recently, researchers have found evidence that filler-gap comprehension begins around 15 months with hallmark idiosyncrasies which models of acquisition must account for [2]. Children seem to find subject-extractions easier to comprehend than object-extractions and wh-relatives easier to comprehend than that-relatives. Both of these asymmetries lessen but are still observable until at least 30-months of age. Further, before 25 months, learners display a bias to assign a unique thematic role to each participant, which leads them to interpret intransitives with conjoined subjects ('John and Mary kradded') as transitives ('John kradded Mary') [3].

Rather than assume that learners must be able to process the complex syntax of a filler-gap dependency, we assume learners simply learn variable word order preferences (which may later be codified as constructions encoded in a grammar). This work uses distributions over sentential positions where possible arguments have been observed in the past to model learner expectations of where arguments can appear relative to a verb. The model assumes that children can identify nouns and verbs and are biased to assign a unique role to each noun. The model is evaluated on role-labelled portions of the CHILDES corpus [1].

Results

First, we compare our model to the previous role-labelling acquisition system [1] and find that our model more closely reflects the performance of children at agent identification. Further, though the models share basic assumptions, analysis of the previous system suggests that it would not perform well at filler-gap comprehension.

Before training, our model assigns roles with moderate accuracy (60% overall, 55% filler-gap) while reflecting the subject/object and that/wh- comprehension asymmetries. The model also reliably obtains the mistaken interpretation of conjoined subjects. That is, the model's performance is characteristic of a 15-month old learner while achieving decent overall performance.

After training on (automatic) part-of-speech tagged child-directed speech, the model achieves a significant improvement in role labelling accuracy (61% overall, 61% filler-gap, both $p<.01$). Overall, the large improvement in object-extraction comprehension (33-48%, $p<.01$) is mitigated by the relative rarity of filler-gap dependencies and by a slight decrease in subject-extraction accuracy (74-72%, $p=.02$) caused by conflict with the newly-formed extracted object position. The model still reflects the previous asymmetric comprehension patterns, but it no longer makes the conjoined subject error. In short, the model now behaves more similarly to a 30-month old learner. Finally, we show that our model is robust to a wide assortment of initializations and training noise.

References

[1] Michael Connor, Yael Gertner, Cynthia Fisher, and Dan Roth. Starting from scratch in semantic role labelling. In *Proceedings of ACL 2010*, 2010.

[2] Annie Gagliardi, Tara M. Mease, and Jeffrey Lidz. Discontinuous development in the acquisition of filler-gap dependencies. Unpublished manuscript: http://www.people.fas.harvard.edu/∼gagliardi, 2014.

[3] Yael Gertner and Cynthia Fisher. Predicted errors in early verb learning. In *Proc. 31st Annual Boston University Conference on Language Development*, 2006.

Poster #1014

Online comprehension of object wh-questions: Eye-tracking evidence against syntactic gap filling

Olga Dragoy, Maria Ivanova, Anna Laurinavichyute (National Research University Higher School of Economics), Anastasia Ulicheva (University of Hong Kong), Svetlana Kuptsova & Lidia Petrova (Center for Speech Pathology and Neurorehabilitation)
odragoy@hse.ru

Available eye-tracking data have been seen as providing support for on-line gap filling during comprehension of object wh-questions. Based on the reactivation of the object NP referent (e.g., *the girl*) at the verb position in *Who did the boy kiss that day at school?,* it was argued that listeners signal a post-verbal trace of wh-movement, and thus use syntactic dependencies to identify the wh-filler referent (Dickey, Choy, & Thompson, 2007). In English, however, the subject NP (*the boy*) must be overtly mentioned in object wh-questions, which makes virtually impossible to identify the reason of the object NP reactivation at the verb. It could be driven either by the syntactic trace of the displaced constituent or by a contextually oriented processing strategy: an overt subject NP cannot be associated with a wh-filler, thus the referential choice is made in favor of another referent involved in an action.

Exploiting flexible word order and case marking of Russian, we carried out an eye-tracking-while-listening experiment to disentangle the effects of syntax and context. 36 native Russian speakers listened to 20 experimental and 20 filler stories presented in a quasi-randomized order. The first three sentences of an experimental story introduced a transitive action, the two critical referents involved in it and two distractor referents: *One day the girl and the boy were walking at school. Suddenly, the boy kissed the girl. The teacher was very surprised*. The story was finalized by a wh-question in either of the two conditions:

(1)	*Kogo*	*malchik*	*potseloval*	*v shkole?*	
	who-ACC	boy-NOM	kissed	at school	(‘*Who did the boy kiss at school?*’)
(2)	*Kto*	*devochku*	*potseloval*	*v shkole?*	
	who-NOM	girl-ACC	kissed	at school	(‘*Who kissed the girl at school?*’)

Participants’ eye movements were recorded as they viewed panels with four drawings depicting referents of a story (two critical – *the boy* and *the girl*, and two distractors – *the teacher* and *the school*). The task required to answer the question at the end by fixating the relevant picture.

Condition (1) was used to replicate previous English results; the wh-word in it referred to the object NP (*the girl*) displaced from the post-verbal position. Thus, both syntax and context drove the same effect: reactivation of the object NP referent at the verb. In contrast, object scrambling in Russian made possible condition (2), in which the displaced constituent was again the object NP (*the girl*), but the wh-filler referred to the subject NP (*the boy*). If the resolution of syntactic dependencies is indeed at the core of object wh-question processing, reactivation of the object NP referent at the verb should be expected. Alternatively, if the processing is context dependent, the overt object NP preceding the verb precludes the possibility that it could be the referent of the wh-filler, and the reactivation of the subject NP referent should follow.

In line with previous findings on English, in condition (1) the proportion of fixations on the object NP referent (*the girl*) was significantly greater than on the subject NP referent (*the boy*) starting from the verb region, with the difference increasing over time. In condition (2), contrary to the syntactically oriented prediction, the proportion of fixations on the object NP referent at the verb was less than on the subject NP referent, and that difference increased over time as well. These results show that identification of the referent associated with the wh-filler is not syntactically constrained, but it is contextually fed: as soon as the context provides enough cues to exclude referents that cannot be linked to the wh-filler, the remaining referent is associated with it. This interpretation is consistent with the view that the wh-filler triggers an active search for its referent, which takes into account other overt referents and even involves an anticipatory mechanism (Sussman & Sedivy, 2003).

Processing filler-gap dependencies in L2: Evidence for the use of subcategorization information

Chie Nakamura (The University of Tokyo), Manabu Arai (The University of Tokyo), Yasunari Harada (Waseda University), Yuki Hirose (The University of Tokyo)
chienak@phiz.c.u-tokyo.ac.jp

Many previous studies in English demonstrated an influence of verb subcategorization information in processing structurally ambiguous sentences (e.g., Stowe et al., 1991), providing evidence for the use of lexically specific information in parsing. However, there is some evidence that L2 learners initially adopt the direct object analysis with an intransitive verb incorrectly in reading a sentence such as *While the audience cried the actor rested behind the curtain*, suggesting that they initially ignore subcategorization information in building the initial syntactic analysis (Nakamura et al., 2013). The finding suggests that L2 learners deterministically adopt the direct object analysis whenever they encounter an NP following a verb independently of verb's lexico-syntactic information. This contrasts with previous studies on L1 processing that demonstrated an influence of verb's structural frequency information or subcategorization preference (e.g., Staub, 2007) and may suggest that lexically-specific syntactic information does not influence structural decisions in L2 parsing. Alternatively, another possibility is that such information in L2 learners is less reliable and thus tends to be overridden by a presumably more influential general principle that an NP directly following a verb should be analyzed as verb's direct object.

To explore this, the current study tested the unbounded dependency structure such as (1) with Japanese EFL learners and examined whether they would nevertheless adopt the direct object analysis by ignoring verb's subcategorization information. The verb was either intransitive (1a; *complain*) or optionally transitive (1b,c; *teach*). The gap-filling NP analysis was either ungrammatical (1a; *complained the building*), semantically implausible (1b; *taught the building*), or plausible (1c; *taught the class*). If the absence of an effect of subcategorization information in previous research was due to the L2 learners' tendency to override lexically-specific information in face of the VERB+NP sequence, the direct object analysis should occur only in (1b, c) but not in (1a). Also, the cost for the semantically implausible direct object analysis should be observed in (1b).

Forty-six Japanese EFL learners were tested using a phrase-by-phrase self-paced reading paradigm. The reading time in Region 6 was marginally longer in (1b) than in (1c) (p=0.065), most likely reflecting a spill-over effect of the implausible direct object analysis from Region 5. The reading time in Region 7 was significantly longer in (1c) than in (1a) (p=0.044). This most likely reflects a spill-over effect of the reanalysis cost from Region 6 where the sentence structure was disambiguated. The results indicate that Japanese EFL learners adopted the direct object analysis only with optionally transitive verbs but not with intransitive verbs, providing evidence for the use of verb subcategorization information in L2 parsing. The results will be discussed with reference to the results for native English speakers.

	Region1	Region2	Region3	Region4	Region5	Region6	Region7
(1a)	That's	the building	that	the lecturer	complained	about	during the semester.
(1b)	That's	the building	that	the lecturer	taught	about	during the semester.
(1c)	That's	the class	that	the lecturer	taught	about	during the semester.

Interference effects of pre-verbal NPs on sentence processing in Japanese

Hajime Ono (Kinki University), Miki Obata (Tokyo University of Science) & Noriaki Yusa (Miyagi Gakuin Women's University)
ono.hajime.home@gmail.com

In a head-final language like Japanese, one may argue that pre-verbal NPs encode a rich source of information for the upcoming predicates. Specifically, it has been argued that a sequence of NP-TOP and NP-NOM triggers a clause boundary insertion before the NP-NOM (Kamide & Mitchell, 1999; Miyamoto, 2002; Aoshima et al., 2004), but the exact syntactic status of the inserted clause remains underspecified. Given the recent suggestion that the syntactic similarities among the encoded words in working memory have a great impact on sentence parsing (Lewis, et al. 2005, 2006; Vasishth & Lewis 2006, 2008; Van Dyke & Lewis, 2003), we conducted a self-paced reading experiment in order to examine what syntactic information is encoded in pre-verbal NPs, using a similarity-based interference effect as an index. We found that pre-verbal NPs encode the full-scale subcategorization information for the upcoming verbs, and that the parser builds a complement clause structure even before seeing the matrix verb.

Twenty-four sets of sentences were prepared in a 2×2 factorial design (A-D), manipulating CLAUSE POSITION (Embedded vs. Pre-posed) and EMBEDDED V-TYPE (NP-V vs. NP/CP-V). The NP-V conditions (A) had the embedded verb that takes only an NP complement (*tsuikyuusuru* 'pursue'), whereas in the NP/CP-V conditions (B), the embedded verb takes either an NP or CP as its complement (*yooseesuru* 'request').

Given the cue-based retrieval approach, upon encountering the embedded verb (in boldface, see below), the parser retrieves the embedded subject [*teachers-NOM*] to establish a subject-verb dependency. In the [EMBEDDED, NP/CP-V] condition (B), if the subcategorization information of the matrix verb (i.e., [matrix V = CP-V]) has been included in the encoding information of the matrix subject [*parents*-TOP], there will be a proactive interference effect due to the partial matches of the retrieval cues of the embedded verb (i.e., NP/CP-V), resulting in a slowdown as compared to the [PRE-POSED, NP/CP-V] condition (D), where there is no such competitor. Furthermore, there should be no interference effect between the two NP-V conditions (A vs. C) because there is no overlap with respect to the subcategorization information.

We found that, at the spillover region ('quietly'), there was an interaction between the two factors ($p<0.05$). Planned pairwise comparisons revealed that there was a reliable difference between the two NP/CP-V conditions (B was reliably slower than D; $p<0.01$) whereas there was no such difference between the two NP-V conditions (A vs. C; ns). We take the reliable difference between the two NP/CP-V conditions as indicating the interference effect. The current results also suggest that the particular sequence of NPs (TOP + NOM) should lead the parser to hypothesize that the embedded clause be a complement CP clause. Taken together, our results strongly suggest that the pre-verbal NP (NP-TOP) encodes information concerning not only its grammatical function and structural position (given the case-marking information), but also the subcategorization information of the expected verb, probably emerging from the structural prediction triggered by the clause boundary insertion. Finally, our results show that the previously encoded information may be updated upon the arrival of further input.

Sample materials

A. Embedded – NP-V [*'pursue'*]
parents-TOP [teachers-NOM principle-GEN fraud-ACC **pursued**-C] quietly talked.about

B. Embedded – NP/CP-V [*'request'*]
parents-TOP [teachers-NOM principle-DAT next.term-ACC **requested**-C] quietly talked.about

C. Pre-posed – NP-V [*'pursue'*]
[teachers-NOM principle-GEN fraud-ACC **pursued**-C] parents-TOP quietly talked.about

D. Pre-posed – NP/CP-V [*'request'*]
[teachers-NOM principle-DAT next.term-ACC **requested**-C] parents-TOP quietly talked.about

The effect of L1 syntax on L2 sentence processing: A self-paced reading study with L2 learners of Chinese

Yun Yao (University of Illinois at Urbana-Champaign)
yunyao1@illinois.edu

In the field of SLA, an essential question that remains unresolved is whether and how L1 affects L2 processing. The Shallow Structure Hypothesis (Clahsen & Felser, 2006) argues against L1 transfer effects on L2 sentence processing; the non-native-like parsing decisions L2 learners make are due to their shallow, less detailed language representations and their inability to use syntax-based processing strategies rather than L1 transfer effects. However, other studies (e.g., Frenck-Mestre & Pynte, 1997) have found that L1 syntax influences L2 processing. These studies suggest that learners apply L1 processing strategies when parsing their L2 and when such L1 strategies do not match the processing of their L2, it results in non-native-like parsing decisions. Therefore, more research is needed to understand how learners' L1 affects the way they process L2 input. In addition, previous studies investigating L1 transfer effects focused on the L2 speakers of Indo-European languages and few have examined the L2 processing of other types of languages. Thus, comparisons between Chinese native speakers and L2 learners' sentence processing strategies can add a new perspective in answering the question of whether and how L1 interacts with L2 processing.

The present study investigated how a temporarily ambiguous construction "Verb NP1 de NP2" was processed by Chinese native speakers and L2 learners from different L1s. The construction is temporarily ambiguous between a relative clause (RC) and a verb object (VO) structure. For example, zhǎodào lǎoshī de xuésheng "find teacher de student" can be interpreted as either 1) "the student who finds the teacher" or 2) "to find the teacher's student." Such ambiguity is mainly caused by the lexical ambiguity of de, which can be interpreted as either an RC marker (as in 1) or a possessive marker (as in 2). The Tuning Hypothesis (Mitchell et al., 1995), the Predicate Proximity principle (Gibson et al., 1996), theories that emphasize thematic role interpretation (Zhang et al., 2000), and empty category interpretation (Ng & Fodor, 2011) all predict that native speakers will prefer the RC interpretation, whereas the Late Closure Strategy (Frazier, 1979) predicts that they will prefer the VO interpretation. For L2 learners, while the Shallow Structure Hypothesis predicts that they will prefer the syntactically simpler VO analysis regardless of their L1s, it is alternatively possible that learners from different L1s will make different parsing decisions.

In Experiment 1, we conducted a pencil-and-paper questionnaire, in which 24 Chinese native speakers, 17 L2 learners whose L1s were head-initial languages (e.g., English), and 16 L2 learners whose L1s were head-final languages (e.g., Japanese) were instructed to rate 15 ambiguous "Verb NP1 de NP2" constructions on a one-to-seven scale. The results showed that the Chinese native speakers strongly preferred the RC interpretation, while the L2 learners did not have a preference for either interpretation. Further analyses showed that while the L2 learners from head-initial languages did not have a preference, the learners speaking head-final languages strongly preferred the VO analysis. Experiment 2 was self-paced reading task, in which the same participants from Experiment 1 read 120 experimental sentences containing "Verb NP1 de NP2" constructions and 75 fillers in a moving window. A 2 (structure) × 2 (ambiguity) design was used; there were four conditions in each experimental sentence set: 1) ambiguous, forced RC analysis, 2) unambiguous, forced RC analysis, 3) ambiguous, forced VO analysis, and 4) unambiguous, forced VO analysis. Results indicated that both the native speakers and the learners from head-final L1s preferred the RC analysis, whereas the learners from head-initial L1s did not show a preference. Taken together, these findings suggest that the Chinese L2 learners whose L1s resembled the target language made native-like parsing decisions when processing Chinese sentences online, whereas those whose L1s differed in headedness seemed to resolve syntactic ambiguity in a non-native-like way.

Long-distance attraction effects in subject-verb agreement processing

Nathan Eversole & Jeffrey Witzel (University of Texas at Arlington)
jeffrey.witzel@uta.edu

A core question in psycholinguistics is how subject-verb agreement (SVA) is computed. One recent proposal is that memory retrieval processes play a key role in SVA during sentence comprehension (Wagers et al., 2009).This model holds that when an agreeing verb (e.g., *was/were*) is encountered, a search is initiated through the memory representation of the sentence for a noun phrase (NP) with matching agreement features. When the search mechanism finds a controlling subject with matching features, the search ends successfully; however, in instances of a mismatch with this subject, the mechanism may incorrectly choose a grammatically inaccessible NP that nevertheless has some of the required agreement features.

A key finding in support of this model is the illusion of grammaticality for sentences like **The musicians that the reviewer praise won a Grammy* (Wagers et al., 2009). On this proposal, the illusion occurs because the plural relative clause (RC) head allows for long-distance agreement attraction -- a finding that "predictive" models of SVA processing (Eberhard et al., 2005; Nicol et al., 1997) have difficulty accounting for. An important question then is whether this attraction effect is observed for more salient agreement targets, such as free auxiliary verbs (*was/were*), compared to inflected main verbs (*praise-s/-Ø).* If this effect reflects core properties of SVA processing, it should be observed regardless of the form of the agreement target.

The present study tested this prediction by examining self-paced reading times (RTs) on sentences as in the examples below. The sentences varied in terms of the number of the attractor NP (the RC head; *musician(s)*) and grammaticality. As in Wagers et al. (2009), EX1 (*N*=40) tested sentences with inflected main verbs (*praise(s)*) as agreement targets. EX2 (*N*=40) tested the same sentences, but with auxiliary verbs (*was/were praising*). EX3 (*N*=40) also used an auxiliary verb, but the verb complex was held constant (*were praising*) and grammaticality was manipulated by adjusting the number of the RC subject (*reviewer(s)*). Ungrammatical sentences with singular attractors were predicted to have longer RTs at and/or immediately after the agreement target. Ungrammatical sentences with plural attractors, however, were predicted to have attenuated processing difficulty in these regions.

EX1 showed an interaction three words after the main verb indicating an effect of ungrammaticality only for singular attractor sentences. In EX2 and EX3, longer RTs were found for ungrammatical sentences at and after the auxiliary, but these effects were significant only for singular attractor sentences under planned comparisons. These findings indicate long-distance agreement attraction effects for both main verbs with bound inflection and free agreeing verb forms and thus provide additional support for a memory retrieval model of SVA processing.

	EX 1	EX 2	
Singular Attractor/Grammatical			
The musician that the reviewer	praises	was praising	so highly won the prestigious award.
Singular Attractor/Ungrammatical			
*The musician that the reviewer	praise	were praising	so highly won the prestigious award.
Plural Attractor/Grammatical			
The musicians that the reviewer	praises	was praising	so highly won the prestigious award.
Plural Attractor/Ungrammatical			
*The musicians that the reviewer	praise	were praising	so highly won the prestigious award.

	EX 3	
Singular Attractor/Grammatical		
The musician that the reviewers	were praising	so highly won the prestigious award.
Singular Attractor/Ungrammatical		
*The musician that the reviewer	were praising	so highly won the prestigious award.
Plural Attractor/Grammatical		
The musicians that the reviewers	were praising	so highly won the prestigious award.
Plural Attractor/Ungrammatical		
*The musicians that the reviewer	were praising	so highly won the prestigious award.

An effect of verb repetition in the production of head-final passive construction

Lan Jin, Ying Deng, Manabu Arai, and Yuki Hirose (The University of Tokyo)
jinlan@phiz.c.u-tokyo.ac.jp

The phenomenon of structural priming has been documented across utterances without any overlap of lexical content, suggesting that the representations of syntactic structures are partly lexically independent. At the same time, many studies also showed that an effect of structural priming is enhanced if the verb is repeated in a subsequent utterance (lexical boost, Pickering & Branigan, 1998), suggesting that syntactic representations are partly associated with the verb. However, most results are based on head-initial languages and relatively little is known about the representations of argument structure in head-final languages. The current study investigates whether argument structure is represented in a similar way in head-final languages by testing the influence of verb overlap between prime and target sentences in Japanese. Althoguh there is some evidence for lexical boost of a priming effect in Japanese (Yamashita et al., 2005), the effect was observed with the repetition of verb morphology in the transitive structures. Our study therefore investigated a priming effect with the Japanese active and passive structures, which do not share the verb morphology. Using a picture description paradigm, we examined whether the boost of passive priming effect could occur in the absence of overlapping post-verbal morphology (i.e., *-rare* for passives). Since the verb morphology differs between actives and passives, our design allows us to distinguish between priming effects caused by verb repetition and those caused by repetition of verbal morphology.

In Experiment 1 (n=28), the verbs were not repeated between primes and targets. The eighteen items consisted of two types of prime sentences (1a, b), and a target picture, which involves two human entities for both agent and patient. After reading the prime sentence aloud, participants were asked to describe the target picture with a designated verb that was presented in its citation form prior to the presentation of the target picture. We observed an effect of priming (p<.01): more passive targets were produced following passive primes (12%) than following active primes (6%). In Experiment 2 (n=28), the verb was repeated between primes (2a, b) and targets. We found a strong effect of priming (p<.001): more passive targets were produced following passive primes (39%) than following active primes (1%). A combined analysis of the data from the two experiments showed a significant interaction between verb repetition and prime type (p<.001), demonstrating that the magnitude of priming was greater when the verb was repeated (Experiment 2, 38%) than when it was not (Experiment 1, 6%), demonstrating a large lexical boost effect (32%).

The current study showed clear evidence for both lexically independent and lexically dependent priming effects. The comparison of the results in the two experiments revealed a substantial lexical boost effect, which appears comparable with previous studies in head-initial languages. This finding suggests that the representation of the passive structure is associated with individual verbs and accessed through the verb even in a language in which the verb is always positioned in the sentence-final position.

Examples

(1)	a. hiroinn-ga	akuyaku-o	karakatta	(The heroine teased the villain.)
	heroine-AGENT	villain-PATIENT	tease-PAST	
	b. akuyaku-ga	hiroinn-ni	karakawareta	(The villain was teased by the heroine.)
	villain-PATIENT	heroine-by AGENT	tease-PASSIVE-PAST	
(2)	a. hiroinn-ga	akuyaku-o	yattsuketa	(The heroine beat the villain.)
	heroine-AGENT	villain-PATIENT	beat-PAST	
	b. akuyaku-ga	hiroinn-ni	yattsukerareta	(The villain was beaten by the heroine.)
	villain-PATIENT	heroine-by AGENT	beat-PASSIVE-PAST	

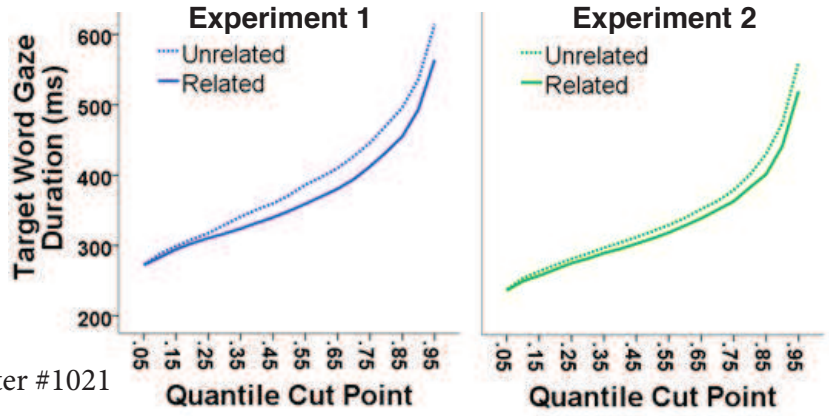
It takes time to prime: Semantic priming in the ocular lexical decision task

Renske S. Hoedemaker & Peter C. Gordon (University of North Carolina at Chapel Hill)
hoedemak@live.unc.edu

Although studies of the recognition of words during sentence processing and when presented in isolation have yielded highly consistent evidence about many facets of word recognition, semantic priming is a notable exception. It is very robust in tasks involving the recognition of words in isolation such as lexical decision tasks (LDT)[1], while quite limited during text reading[2]. Ex-Gaussian distribution fits of manual LDTs have shown relatedness effects across the RT distribution (reflected by changes in μ) suggesting an encoding-based priming mechanism[3] by which a target word's meaning is partially pre-activated as a result of processing a related prime. We evaluate the contributions of response mode and task goals to semantic priming by replacing the manual response mode typically used in LDT with an eye-movement response through a sequence of three words. Experimental prime-target pairs appeared in the initial and middle position of the triplet. This *ocular* LDT combines the explicit control of task goals found in LDTs with the highly practiced ocular response used in reading text.

In Experiment 1, forward saccades indicated an affirmative LD on each word in the triplet. Although gaze durations (M = 426 ms) were much shorter than manual LDs for the same words in the English Lexicon Project[4] (M = 605 ms), t = 59.9 $p < .001$, we observed a reliable semantic priming effect, $t_1 = 5.2$, $p < .001$, $t_2 = 4.9$, $p < .001$. In contrast to findings in manual LDTs[3], semantic priming in the ocular LDT was most pronounced for slower responses, with ex-Gaussian fits showing a strong effect on τ, t = 2.6, $p < .02$, while the effect on μ was only marginal, t = 1.8, $p = .08$. This pattern suggests that ocular LDs reflect a task-based recruitment of prime information on trials with more effortful target resolution. In Experiment 2, LD responses were delayed, so that subjects indicated the total number of nonwords in each triplet after all three letter strings had been read. Word reading times showed a semantic priming effect even though the eye movements were not directly related to an explicit task goal, $t_1 = 4.1$, $p < .001$, $t_2 = 3.1$, $p < .01$. Average word reading times (M = 372) were shorter than those in Experiment 1, t = 20.1, $p < .001$, and Ex-Gaussian distribution fits revealed an even greater concentration of the semantic priming effect in estimates of τ, t = 2.18, $p < .05$, while the effect on μ was not significant, t = 1.2, $p = .23$.

These findings do not support the interpretation that semantic priming results from an encoding-based pre-activation process since priming affected slow but not fast responses. This dependence of priming on trial time/difficulty may be obscured by the high response-time floor in studies that use manual LDTs. Semantic relatedness effects during sentence reading may be fleeting because words are rapidly and easily recognized in this highly practiced task. Even though first-pass reading measures are considered especially sensitive measures of lexical encoding[5], the ocular LDT suggests that semantic priming effects in isolated word recognition result primarily from task-related, decision-based processes.



[1] Meyer & Schvaneveldt (1971) *JEP* [2] Camblin et al. (2007) *JML* [3] Balota et al. (2008) *JML*. [4] Balota et al. (2007) *BRM* [5] Rayner (1998) *Psych Bulletin.*

Computing the structure of questions: Evidence from online sentence processing

Hadas Kotek and Martin Hackl (MIT)
hkotek@mit.edu

Summary: Recent theories of interrogative syntax/semantics adopt two strategies for the interpretation of *wh*-in-situ: *Covert movement* and *in-situ computation*. The covert movement strategy is traditionally assumed to be all-or-nothing: *wh* covertly moves to C at Logical From, or else is interpreted in its base position. We test the predictions made by these approaches for the processing of in-situ *wh*-phrases and report two major findings: (a) *wh*-phrases behave like traditional quantifiers (*every*) and unlike non-quantificational determiners (*the*) in their ability to facilitate Antecedent Contained Ellipsis (ACE) in online sentence processing. (b) *wh*-phrases, but not traditional quantifiers, are sensitive to the presence of *focus interveners* in the structure. We argue that our findings are explained if *wh*-phrases require *both* covert movement and in-situ computation for their interpretation: "in-situ" *wh*-phrases undergo a *small* covert movement step and can then be interpreted *in-situ* in the lowest position where they are interpretable.

Real time processing of ACE: *Wh*-phrases can host ACE which, in example (1), can be either ellipsis of a *local* VP_2 (<perform…>) or *non-local* VP_1 (<BE willing to perform…>). ACE requires a constituent containing the ACE-site to take scope at least as high as the antecedent. We use the methodology in Hackl et al. (2012), who show that in online sentence processing, *every* but not *the* can facilitate the resolution of *local ACE*. This is predicted by QR-based theories of quantifier integration, because QR above VP_2 must be assumed by the parser when *every* is encountered to resolve a type-mismatch. This preemptively undoes antecedent containment, making the ellipsis easier to resolve. Non-local ACE is not facilitated because QR is not sufficiently high to undo containment. The structure must be reanalyzed when the ACE-site is encountered, incurring high processing costs. With *the*, no movement is assumed when the determiner is encountered; neither local nor non-local ACE resolution is thus facilitated.

(1) The conductor asked [$_{CP}$ *which* soloist [$_{VP1}$ was willing to [$_{VP2}$ perform…
  a. ***which*/*every/the*** concerto that the brilliant protégé **did**… local ACE
  b. ***which*/*every/the*** concerto that the brilliant protégé **was**… non-local ACE
  …and restructured the rehearsal accordingly.

Exp.1: We used a paradigm like (1) to test if the processing of *wh*-in-situ involves covert movement. Under the *covert movement* approach, the *which*-DP (with the relative clause and ACE-site) moves *non-locally* to C when *wh* is encountered. Ellipsis resolution is relatively easy for both *local* and *non-local* ACE compared to *every*. Under the *in-situ* approach, there is no movement of the *which*-DP. The structure is reanalyzed when the parser encounters the ACE-site. Both *local* and *non-local* ACE incur relatively high processing costs compared to *every*. We find that *which* patterns with *every* and unlike *the*: it is able to facilitate the resolution of *local ACE* but not *non-local ACE* (n=161, linear mixed effects models on logRT two words after the ACE-site, $p<.05$). This is not predicted by either approach to *wh*-in-situ. We propose that it is explained if *wh*-phrases are quantifiers that must QR *locally* when they are parsed.

Exp.2: Our proposal predicts that longer-distance covert movement of a *wh*-phrase can be caused by material that triggers so called *intervention effects* (Beck 2006, a.o). When an *intervener* occurs above the *wh*-phrase, the *wh*-phrase must move above it for interpretation. We show that *which* but not *every* is sensitive to the presence of the intervener *also*. We place *also* in one of two positions: (a) above the lower VP_2 ('was willing to *also* perform…') or (b) above the higher VP_1 ('was *also* willing to perform…'). *Also* is predicted to force covert movement of *which* above it, resulting in an increased domain of ACE facilitation effects. Results show *non-local ACE* is facilitated with (b), but only *local ACE* is facilitated with (a). *Every* is not affected by the presence of the intervener: only *local ACE* is facilitated in both (a) and (b) (n=89, linear mixed effects models on logRT two words after the ACE-site, $p<.05$). This confirms the role of covert movement in the structures under discussion.

Online processing of English Which-questions by children and adults: a visual world study

Carla Contemori (Penn State University) and Theo Marinis (University of Reading)
cuc29@psu.edu

Much work on child sentence processing has demonstrated that children are able to use various linguistic cues to incrementally resolve temporary syntactic ambiguities, but they fail to use syntactic or interpretability cues that arrive later in the sentence. The present study explores whether children incrementally resolve filler-gap dependencies, using English Which-Object questions (O-WH). We compare off-line and on-line processing of O-WH questions (O-WH) and S-WH questions (S-WH), and we manipulate the number features of the subject and object NP within the two question types (O-WH singular-plural: *Which cow are the goats pushing?;* O-WH singular-singular: *Which cow is the goat pushing?*; S-WH plural-singular: *Which cows are pushing the goat?;* S-WH singular-singular: *Which cow is pushing the goat?*). The aim is two-fold. Firstly, we observe if any facilitation emerges in the off-line comprehension of O-WH questions when there is similarity or dissimilarity in terms of number features. Secondly, we use eye-tracking to explore the role of number features in the processing and reanalysis of filler-gap dependencies.

Thirty one English-speaking children (5–7;10) and 21 adults participated in a visual-word paradigm task. Eye-movements were recorded while participants looked at two pictures (target, competitor) and listened to a S-WH or O-WH question. Eye-movements were time-locked to the onset of the auxiliary verb (is/are) and divided into 5 time-windows (TWs) of 400 ms each.

Accuracy scores showed that children were significantly less accurate than adults for all O-WH questions, and they were significantly more accurate in O-WH questions with number mismatch (singular-plural), than match (singular-singular). Eye-tracking data for correctly comprehended sentences suggested that for all Which-questions, adults and children began their interpretation by postulating a subject question. This was shown by a significant higher amount of looks to the incorrect target for O-WH questions compared to S-WH up to 1200 ms after the onset of the auxiliary. Children started revision for O-WH questions significantly later than adults (Fig.1,2), as shown by an interaction between Sentence type (O-WH vs S-WH) and Group (adults and children) from 800 ms to 1200 after the auxiliary. Furthermore, children took longer to revise their interpretation of O-WH singular-singular than O-WH singular-plural, as shown by an interaction between Group, Sentence type and Matching in TW4. The accuracy results show that children are more likely to perform revision accurately when there is a mismatch of number features between the subject and object NP in O-WH questions. The eye-tracking results indicate that (1) revising the initial subject interpretation in O-WH questions takes longer for children compared to adults, and (2) children – but not adults - commit more strongly to a subject interpretation for O-WH with match (singular-singular), than for O-WH with mismatch of number features (singular-plural). Results will be discussed in the context of current psycholinguistic theories of syntactic processing and models of language acquisition.

Fig 1. Proportions of looks to Target and Competitor in O-WH singular-plural

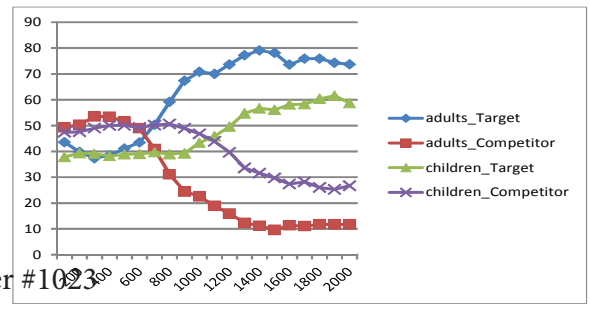
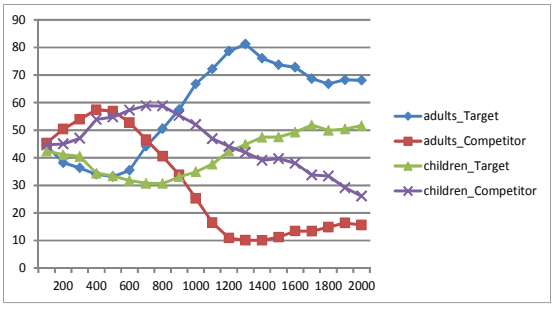


Fig 2. Proportions of looks to Target and Competitor in O-WH singular-singular



A Bayesian meta-analysis of studies on Chinese relative clause processing

Shravan Vasishth (University of Potsdam)
vasishth@uni-potsdam.de

Hsiao and Gibson [1], and more recently Gibson and Wu [2], have presented evidence from self-paced reading which suggests that subject relatives in Chinese are harder to process than object relatives. This finding is important because it confirms a key prediction of working-memory based sentence processing accounts. The explanation for the object-relative (OR) advantage in Chinese under these accounts is that the distance between the gap in the relative clause and the head noun is larger in subject vs object relatives. This leads to greater retrieval difficulty (due to decay and/or interference), leading to longer reading times at the head noun in subject relatives compared to object relatives.

The claim that there is an OR advantage in Chinese is highly contentious because other studies on Chinese RCs present a subject-relative (SR) advantage. The SR advantage can be explained in terms of complexity metrics that rely on structural frequency based information, but other explanations also exist.

A key issue here is: what should we believe about Chinese RC processing given the data so far? Is there an SR advantage or an OR advantage? We address this issue by carrying out a bayesian random effects meta-analysis using the results of 14 published or publicly available Chinese RC studies, all of which compared reading times at the head noun in SRs and ORs, and all of which used animate nouns. Meta-analysis is a well-known methodology in medicine and other areas and is used primarily for evidence synthesis. An important advantage of using the bayesian approach is that we can compute the posterior probability of the coefficient being negative or positive given the data; that is, the posterior probability that Chinese has a subject or object relative advantage given the evidence available. One assumption here is that the difference between SR and OR processing time (the study effect) in each of the studies (which have comparable designs but different items, subjects, sample size, and differences in how and where they were run) comes from a normal distribution with its own mean and standard deviation. The study effects themselves are assumed to come from a normal distribution with an unknown mean and standard deviation, representing the true effect (this unknown true effect is what is estimated in analyses of single studies).

The model yielded a posterior probability of an SR advantage to be 0.79. Thus, to the extent that we want to base our beliefs on empirical evidence (letting the data speak for themselves), the existing data point to a subject-relative advantage in Chinese.

We also systematically investigate how skeptical priors—reflecting different strengths of prior beliefs for an OR advantage—modulate the posterior distribution. This allows us to quantify our uncertainty about the question given the data and given a strong prior belief, which could be justified given theory. More generally, this approach provides a formal tool for quantifying our uncertainty about what the facts are about particular phenomena given the data so far, and allows us to formally update our beliefs as new data emerge. Eventually, agreement between competing beliefs can be reached by incrementally updating the posterior distribution given new data.

[1] F. Hsiao and E. Gibson. Processing relative clauses in Chinese. Cognition, 90:3–27, 2003.
[2] E. Gibson and I. Wu. Processing Chinese relative clauses in context. Language and Cognitive Processes (in press), 2011.

Nonadjacent lexical dependencies in an artificial language prime relative clause attachment biases

Felix Hao Wang, Mythili Menon & Elsi Kaiser (University of Southern California)
wang970@usc.edu

Research in child language development and adult sentence processing has led to intriguing insights about the representation of abstract dependencies in language and other cognitive domains (e.g. Gomez'02, Scheepers et al'11, van de Cavey/Hartsuiker'11). We build on these insights to explore whether *an abstract relation represented through word-level statistical regularities in an artificial language can prime the attachment biases of relative clauses (RCs).* Can adjacent and non-adjacent structures derived from statistics prime the low vs. high attachment preference during the production of relative clauses?

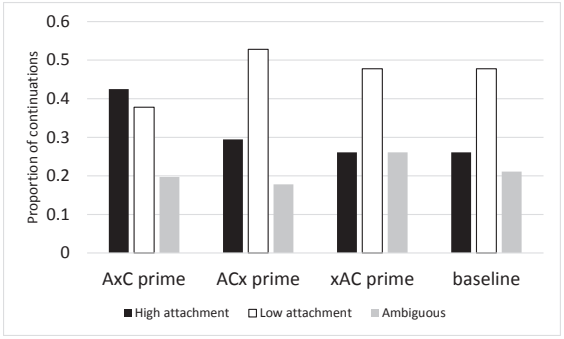
Our study directly manipulated the priming material, with is non-adjacent dependency sequences (A_iXC_i) from Gomez'02. A non-adjacent dependency is a three-word sequence such that the first word uniquely predicts the third word, while the second word can vary. This structure models the linear sequence of high attachment in relative clauses (e.g., *Kevin counted the fans of the singer who were excited*). Our hypothesis is that if abstract relations can trigger syntactic priming, learning the non-adjacent dependency should prime participants to produce more high-attachment relative clause completions than controls.

EXPERIMENT: Participants (n=50) first completed a *training phase* where they heard three-word strings in an artificial language (adapting stimuli from Gomez'02, e.g. *choon blit jub*). Training was followed by a *test phase,* where participants heard strings and indicated whether they had heard those strings during training. During the test phase, participants also wrote completions to sentence fragments, which on critical trials were ambiguous RC fragments (e.g. *Kevin counted the fans of the singer who*). The completed RC can either provide information about *the fans* (high attachment, HA) or *the singer* (low attachment, LA).

On critical trials, the artificial language prime occurred immediately before the RC fragment. In a between-subjects design, we manipulated the dependencies that participants encountered during the training and test phases: One group was trained on non-adjacent dependencies ($\mathbf{A_iXC_i}$), two groups on adjacent dependencies ($\mathbf{A_iC_i}$X, and X$\mathbf{A_iC_i}$), and a fourth baseline group on random three-word orders with no dependencies. Participants' RC completions were analyzed as HA, LA, or ambiguous.

PREDICTIONS: English RC's have a default low-attachment bias (Mitchell/Brysbaert'98). Can this be weakened by nonadjacent primes? If nonadjacent lexical-level dependencies prime dependency formation in RCs, participants should produce more high-attachments after non-adjacent than adjacent sequences (A_iC_iX, and XA_iC_i).

RESULTS/LEARNING. Participants successfully learned non-adjacent dependencies when exposed to them: They performed above chance in test-phase learning questions ($p<0.001$, LME). **RESULTS/PRIMING**: Crucially, we also find priming of RC completions: There are significantly more high-attachment completions in the nonadjacent prime group than the other 3 groups ($p<0.001$, LME), as shown in the figure. People who had learned a non-adjacent dependency in the artificial language were more likely to produce RCs that attach to the non-adjacent (higher) noun, compared to people who learned local dependencies or no dependencies.



IN SUM, our study provides novel evidence that abstract relations represented through newly-learned word-level statistical regularities can prime the attachment biases of RCs, and suggests that the underlying representations are best regarded as highly abstract and finely attuned to statistical regularities in the input.

The Processing and Neural Basis of Argument Structure Composition through Eye-Tracking, Focal Brain-Lesion and fMRI

Sara Sanchez-Alonso[1], Amy Ly[1], David Braze[2], Cheryl Lacadie[1], Todd Constable[1] and Maria Pinango[1] (Yale University[1] and Haskins Laboratories[2])
sara.sanchez.alonso@yale.edu

Light verb constructions (LVCs) such as *The woman **gives** an order*, contrast with *heavy* counterparts like *The woman **types** an order*. While both sentences exhibit the same surface syntax: [NP [V NP]], only in the LVC the argument structure (AS) of the sentence emerges from the object's meaning, *order*. Whereas all analyses agree that LVCs result from the integration of the (underspecified) AS of the light verb [give$_{<x,y,z>}$] and the AS of the object [order$_{<ordered,order,orderee>}$], proposals differ regarding how this integration takes place. Such differences carry in turn distinct processing and cortical implications: Under the ***lexical approach,*** LVs are built and stored in the mental lexicon, similar to idioms[1,2]. Accordingly, LVCs are retrieved as constructions and are highly sensitive to frequency factors. LVs have been shown to be more frequent than their heavy verb (HV) counterparts[3]. Consequently, LVCs should be retrieved faster than non-LVCs. By contrast, the ***compositional approach*** claims that LVCs are formed in real-time by composition of the semantically underspecified verb and the object's AS[4,5]. As a result, when frequency is controlled, LVC composition may be observed as higher cost during online comprehension.

Study 1 tested the underspecification claim**.** An eye-tracking study (n=36, 40Ss/cond, English) compared LVCs and their HV counterparts. Each sentence was preceded by a context (*On the main deck of the boat,*) and followed by either an LVC (*David **gave** a **shout** to Captain McDonald…*) or a HVC (*David **heard** a **shout** from Captain McDonald…*). Both conditions were controlled for frequency. ***Results***. Greater fixation times were observed for HVs at the verb's region. As frequency was controlled, these findings support the underspecification description of LVs previously shown for German[6], and consequently for the compositional approach to LVs since the cost of lexical retrieval of the HV must be greater than the LV's, which is underspecified.

Study 2 tested the compositional claim. 7 Broca's, 4 Wernicke's and 8 matching controls were tested on three conditions using a picture-matching task (40Ss): LV (*La mujer **hace** una **llamada**,* "The woman makes a call"); HV (*La mujer **recibe** una **llamada***, "The woman receives a call") and DarkV (*La mujer **hace** un **círculo***, "The woman makes a circle"). Previous lesion and neuroimaging results associate argument structure composition with Broca's area, and not with Wernicke's[7]. Accordingly, the compositional approach, but not the lexical approach, predicts that only Broca's will show poor performance for LVs. ***Results.*** *Only for LVCs:* Broca's < Wernicke's = Controls (Fig.1).

Figure 1. LVCs in Broca's aphasia

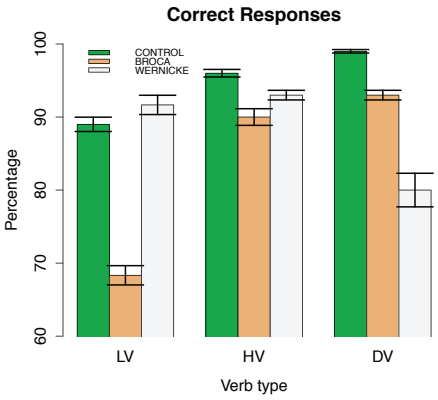
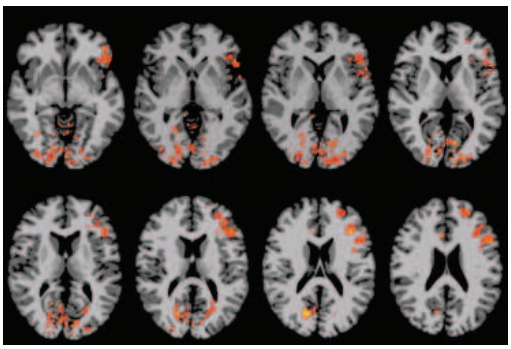


Figure 2. **Brain activations of LV-HV.**



Study 3. An fMRI study (n=26, 50Ss/cond, English) also tested compositionality. ***Results***. LVC composition preferentially recruits the left frontal gyrus (LFG) (Fig.2). These results support the compositional analysis of LVCs. Real-time compositionality of LVCs suggests architecture of the processor whereby semantic composition is taking place partially independently and parallel to syntactic composition.

Turkish-speaking children use verbal morphemes to predict argument structure

Duygu Özge (Harvard University), Deniz Özkan-Göktürk, Berna Arslan-Uzundağ & Aylin Küntay (Koç University), Jesse Snedeker (Harvard University)
dozge@wjh.harvard.edu

Studies of the development of sentence processing have focused largely on languages like English in which verb information and word order are critical for online thematic role assignment. A smaller literature demonstrates that children acquiring head-final languages use case-marking to predict upcoming arguments (Choi & Trueswell, 2010; Özge, Küntay, & Snedeker, 2013). Tagalog children combine the case marking and the causative morpheme to recover the presence of an implicit argument (Trueswell, Kaufman, Hafri, & Lidz, 2012). The present study explores whether children can use verbal morphology alone for thematic role assignment. Specifically, we investigate whether Turkish-speaking children can use relativizing and passivizing morphemes to assign thematic roles to dropped arguments (a context where there no case-marking cues are available).

We conducted two experiments to compare the comprehension of subject- versus object-relativizing morphemes (Exp-1: 1 vs. 2) and subject relativizer versus a combination of passivizer and a subject relativizer (Exp-2: 1 vs. 3). Each experiment tested three age-groups of children (4;0-4;6, 5;6-6;0, 8;0-9;0) using a sentence-referent matching task. Participants were presented with a picture with an animal acting on another animal and pre-recorded questions, like (1)-(3), and were asked to answer the question by touching the correct animal. The correct response is the Agent for (1), it is the Patient for (2) and (3). To eliminate case-marking cues, all arguments were omitted, which is both grammatical and pragmatically appropriate in Turkish.

(1) Hangisi it-en?
which push-SRel
'Which is the one that is pushing it?'

(2) Hangisi it-tiğ-i?
which push-ORel-Poss.3sg
'Which is the one that it is pushing?'

(3) Hangisi it-il-en?
which push-Pass-SRel
'Which is the one that is being pushed?'

For each experiment and for each age group we analyzed whether the type of the verbal morpheme had any effect on the Agent Preference (i.e., # of Agent Responses - # of Patient Responses). In Experiment 1, the number of Agent responses was significantly greater in the Subject-Relatives compared to the Object-Relatives only in the 8-year-old group ($z = -2.54$; $p = .01$). In Experiment 2, the number of Agent responses was significantly greater in the Subject-Relatives compared Passive Condition in all age groups (4-year-olds: $z = -3.08$; $p = .002$; 5-year-olds: $z = -2.06$; $p = .003$; 8-year-olds: $z = -2.9$; $p = .004$).

The study revealed for the first time that verbal morphemes per se can guide thematic role assignment even when there is no accompanying case marking cues. The object-relativizing morpheme leads to more erroneous Agent Responses, probably due to its ambiguous function (i.e., it could mark relativization or complementation). This might also explain the observed preference for subject- versus object-relative clauses in Turkish (Slobin, 1986; Özge, Marinis, & Zeyrek, 2010). The present study also demonstrates that the subject relativizer is acquired earlier than passivizer, which is acquired earlier than object relativizer.

Two Distinct Attraction Profiles in Comprehending Russian Gender Agreement

Natalia Slioussar (Utrecht Institute of Linguistics OTS, St.Petersburg State Univeristy),
Anton Malko & Colin Phillips (UMD)
amalko@umd.edu

Agreement attraction (e.g., *The key to the cabinets are …*) has received much cross-language attention in language production, but little attention in language comprehension, with most evidence drawn from number attraction in English and Romance. Number attraction is an error involving an independent morphological feature that can combine with most any noun, and it has received diverse accounts, that appeal to a mis-represented target noun (feature percolation [1], noisy channel [2]) or retrieval of features of the incorrect noun [3]. Here we report a self-paced reading study on gender attraction effects in Russian comprehension, which contributes to this debate in a number of ways. First, unlike number, gender is an inherent feature of nouns, and thus less amenable to percolation or noisy recall. Indeed, some evidence from production suggests that gender attraction is less robust than number attraction, or even absent [4-6]. We find that Russian masculine (M) agreement yields an identical attraction profile to English number, motivating a retrieval account. Second, previous comprehension evidence, mostly from English number, suggests that attraction has a uniform profile. 'Marked' features, e.g., *plural*, cause facilitatory interference in ungrammatical sentences, but there is little or no interference in grammatical sentences, or from unmarked features. This profile has been used to motivate a retrieval-based account in which a single parallel direct-access mechanism is engaged to retrieve agreement features specifically when the expected agreement form is not encountered. We find that Russian feminine (F) and neuter (N) attractors also induce attraction, but with a rather different profile than for M attractors. This motivates the proposal that attraction engages different retrieval processes for unmarked (M) and marked (F, N) features.

Russian speakers (n=40) read 8-9 word sentences in which a past tense predicate matched or mismatched in gender with a subject noun ('head') or a noun embedded inside a PP complement ('attractor'), following the scheme in (1). The 3-way Russian gender system motivated a 16-condition design: we manipulated grammaticality and attraction (2x2) for each of four combinations of head and attractor noun (F-M, N-M, M-N, M-F). The predicate marked gender on both the copula and the adjective/participle. The head NP was nominative. The preposition selected for accusative nouns, but the attractor NPs were chosen such that their accusative form was identical to the nominative. Targets were embedded among 120 filler items. One third of sentences were followed by forced choice comprehension questions.

(1) NP_{SUBJ} – *P* – $NP_{INTERFERER}$ – *(copula)V – Adj/Part … [4 additional words]*

Results in the sets of conditions with masculine attractors (F-M, N-M) replicate previous findings from number attraction in English and Romance. Reading times immediately after an ungrammatical verb showed facilitation in the presence of an attractor (19ms), but there was no corresponding disruption in grammatical sentences.

In contrast, in the sets of conditions with non-masculine attractors (M-F, M-N) we found that the attractors were uniformly disruptive. Reading times immediately after an ungrammatical verb were even slower when the verb matched the gender of the attractor (29ms), and similar disruption was observed in grammatical sentences (31ms), in cases where the attractor mismatched the gender of the verb. This profile has not been observed in previous studies.

Taken together, we suggest that these findings motivate a modified retrieval-based account of attraction, in order to account for the two distinct attraction profiles. We propose that (i) only heads with marked gender (F/N) initiate predictions, as only these are associated with overt verb agreement; (ii) retrieval is initiated at the verb if there is either no prediction or an unmet prediction; (iii) the retrieval mechanism engages differently with unmarked (M) and marked (F/N) attractors. The parser engages a shallow check for the presence of an M feature, which allows it to mistake structurally inappropriate attractors for grammatical licensors. But the more prominently represented F/N nouns engage more thorough processing, leading to recognition of the structural mismatch and ensuing conflict-induced slowdown.

References
[1] Eberhard et al., PsychRev 2005. [2] Bergen et al., CUNY 2013. [3] Wagers et al., JML 2009.
[4] Lorimor et al., LCP 2008. [5] Vigliocco et al. JML 1995. [6] Vigliocco & Franck, JML 1999.

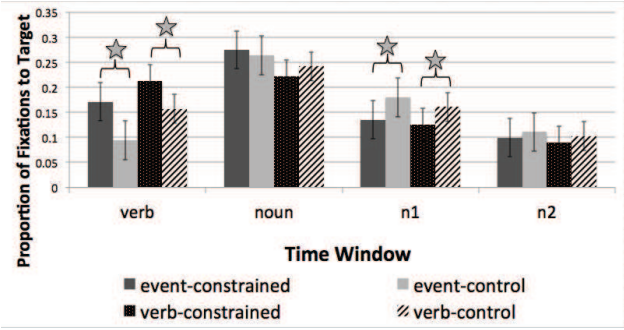
No lexical boost: verb-based information does not facilitate prediction over and above event-based knowledge in the visual world

Evelyn Milburn, Tessa Warren (University of Pittsburgh), & Michael Walsh Dickey (University of Pittsburgh; VA Pittsburgh Medical System)
eam115@pitt.edu

Evidence suggests that comprehenders rapidly predict on the basis of language and event knowledge (e.g., Kamide et al, 2003), but we know little about how these kinds of knowledge interact in driving prediction. We report a visual-world eye-tracking study comparing prediction driven solely by event knowledge derived from natural scenes, to prediction driven by both event knowledge and verb-based information.

College-aged adults (n=36) viewed photographs of natural scenes while listening to sentences (e.g. Staub et al, 2012). There were two critical conditions: verb-constrained sentences wherein the verb's selectional restrictions and the scene constrain likely upcoming objects ("Someone will strum the guitar": *strum* requires a strummable object and the only such object in the scene is a guitar); and event-constrained sentences wherein only the event depicted in the scene constrains likely upcoming objects ("Someone will fling the flowers", presented with a scene of a bride at a wedding). Each constraining sentence was compared to a corresponding non-constraining control sentence. Cloze norms on sentences alone verified that the target was highly expected only in the verb-constrained condition. In contrast, with concurrent presentation of the sentence and image, the target was similarly highly expected in both constraining conditions, which differed only in the presence or absence of prediction-driving verb information.

Participants anticipated upcoming direct objects similarly in both constrained conditions. In a window extending from 200 ms post verb onset to 200 ms post noun onset (labeled 'verb' in graph), there were main effects of: (1) constraint, with more anticipatory looks to the target in the constrained vs. control conditions ($F_1(1,14)=18.3$, $p<.05$; $F_2(1,35)=23.4$, $p<.05$), and (2) sentence type, with more looks to the target in the verb-constrained and verb-control sentences than in event-constrained and event-control sentences ($F_1(1,14)=7.9$, $p<.05$; $F_2(1,35)=15.3$, $p<.05$). (The sentence-type effect reflects between-item differences.) Critically, there was no interaction, and temporally fine-grained plots also did not indicate an interaction. A 500 ms window beginning 200 ms post noun offset (labeled 'n1' in graph) showed a reversed main effect of constraint ($F_1(1,14)=9.9$, $p<.05$; $F_2(1, 35)=8.5$, $p<.05$).



These results suggest that the primary driver of prediction here is event-based knowledge constrained tightly by the scene and loosely by the verb. There is no evidence that the addition of strong but redundant verb-based constraints influences prediction, at least as measured by anticipatory gazes in the visual world. This finding is consistent with previous research suggesting that event-related knowledge is used very early during language comprehension (McRae & Matsuki, 2009). It provides no evidence that strong verb-related knowledge (like selectional restrictions) might generate earlier or more advantaged predictions than scene-based event knowledge—a possibility that would be consistent with literature on selectional restriction comprehension (e.g., Paczynski & Kuperberg, 2012; Warren & McConnell, 2007).

Event representations in collective and distributive readings: An on-line study

Garrett Smith (University of Connecticut)
garrett.smith@uconn.edu

The meaning of collective and distributive sentences can be characterized in terms of the event structures they describe [2]. A collective sentence like (1) "John and Mary carried a box together," is true iff there was a box-carrying event in which John and Mary together were the agent. Similarly, the mental representation constructed upon reading (1) should involve a single representation of that event. But a distributive sentence like (2) "John and Mary each carried a box," is true iff there was a complex event consisting of two box-carrying sub-events with John as agent of one and Mary the agent of the other [2]. The experiment reported here provides evidence that distributive sentences are mentally represented at the finer-grained level of multiple sub-events.

The present experiment uses a modified self-paced reading paradigm where participants read sentences presented centrally in one- or two-word groups and enter whether one or two words are on screen (1/2-task) when a word group is a different color [3, 4]. Critical word groups contain one word. In [3], 1/2-task reaction times were longer in distributive sentences than in collective sentences on the word group containing a singular direct object noun. The singular-marked direct object seemed to be mentally represented with one noun token per sub-event, and this implicit plurality interfered with the task of entering that only one word was on screen. In [4], the 1/2-task was done using intransitive, inherently distributive verbs to investigate whether distributive sentences with plural subjects have mental event representations at the sub-event level. The interference effect between the plurality of sub-events and the 1/2-task was only found when the plural subject consisted of a conjoined NP ("the machine and the computer"), but not when it was a simple plural definite description ("the machines"). The question arises as to what the cause of the interference effect is: is it a surface-level phenomenon due to the separate introduction of referents, or is it due to multiple sub-event representations?

Using the same 1/2-task, 28 native speakers of German read collective and distributive sentences in German. Pronominal subject + adverb combinations were used to set the collective (*sie zusammen (they together)*) or distributive (*beide nacheinander (both one after the other)*) reading. Also, all sentences contained only singular direct objects (in contrast to [3], which used both singular and plural direct objects): (3) *Die Umzugshelfer / Erik / und Jan / haben / eine schwere / Kiste gepackt, / die / {sie zusammen* or *beide nacheinander} / kurz / tragen*. (The movers / Erik / and Jan / have / a heavy / box packed, / which / {they together *or* both one after the other} / a short time / carry.) (Slashes mark word groups. The grouping of words into word groups varied by item.) Participants performed the 1/2-task on the final verb (*tragen*).

A linear mixed effects model with crossed random effects for participants and items was fitted to the 1/2-task reaction time data. Absolute t-values greater than 2 were considered significant [1]. The model showed that reaction times in the distributive condition were longer than in the collective condition (t = -4.047), indicating interference between multiple sub-event representations and the 1/2-task, even though the individuals in the subject were not introduced separately. These results suggest that the interference effect in [4] is indeed due to plural event representations, not overtly separate referents, and strengthen the evidence that understanding a distributive sentence involves constructing a mental representation with multiple sub-events.

References

[1] R. Baayen, D. Davidson, and D. Bates. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59:390--412, 2008.
[2] F. Landman. Plurality. In S. Lappin, editor, *The handbook of contemporary semantic theory*, pages 425--457. Blackwell, 1996.
[3] N. D. Patson and T. Warren. Evidence for distributivity effects in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3):782--789, 2010.
[4] N. D. Patson and T. Warren. Event distributivity and plurality. Poster at AMLaP 2011 in Paris, France, 2011.

Comprehenders mentally represent some aspects of plural entities

Nikole D. Patson (Ohio State University)
Huffman.689@osu.edu

A number of studies have suggested that comprehenders build detailed, image-like conceptual representations of singular entities in sentences (e.g., Zwaan et al., 2002). Recently, Patson, George, and Warren (2013) argued that the conceptual representations of plural entities do not contain the same level of detail as singular entities. Using a NP/picture-matching paradigm (Zwaan et al., 2002) Patson et al. found that when the sentence contained a singular definite description (e.g., *apple*), participants were faster to respond to a picture of a single object than a picture of multiple objects. However, when the sentence contained a plural definite description (e.g., *apples*) there was no time advantage for a picture of a small set of objects compared to a picture of a single object. This finding suggests that plural definite descriptions are left conceptually underspecified. However, one alternative is that the conceptual representations of plural definite descriptions contain detailed information about group-level properties at the expense of properties of the individual entities that make up the group. The two experiments reported here use the same NP/picture matching paradigm to investigate this hypothesis.

In Experiment 1 sentences described sets of objects that are physically close together (1a) versus distributed in space (1b) and pictures either contained sets of closely grouped objects or a set of scattered objects. After reading a sentence, 60 participants judged whether or not a picture was of a set of objects that were mentioned in the sentence. In both experiments, participants were instructed to base their judgments on object identity and not features of the set. Thus, the correct answer for all experimental items was “yes”. Picture judgment times (see Table 1) showed a main effect of picture match such that judgment times were fastest when the picture matched the implied distribution of the objects.

In Experiment 2 (N = 48), sentences contained plurals that implied large, uncountable set sizes (2a) versus small, countable set sizes (2b) and pictures contained a small, countable number of objects or a large, uncountable number of objects. Picture judgment times (see Table 1) showed an interaction between linguistically implied set size and picture condition. When the sentence implied a small set of objects, participants were faster to respond to a picture of a countable set of items than to a picture of an uncountable set. However, when the sentence implied a large set of items there was no judgment time difference between the two picture conditions.

These findings suggest that plural definite descriptions are not always conceptually underspecified. Instead, when a plural definite description refers to a countable set of objects, its conceptual representation contains details about the group features. However, when the plural definite description refers to an uncountable number of entities, its conceptual representation is left underspecified (Johnson-Laird, 1983). This is consistent with proposals in which conceptual representations are detailed and potentially grounded in experience (Barsalou, 1999).

Table 1. Items and Picture Judgment Times (correct responses only) for Expts 1 and 2.

	Match	Mismatch
1a. The child stacked the blocks.	1298 ms	1565 ms
1b. The child knocked over the blocks.	1322 ms	1666 ms
2a. The family stood in the patch full of pumpkins.	1586 ms	1520 ms
2b. The family decorated their porch with pumpkins.	1371 ms	1982 ms

Misinterpretation in agreement and agreement attraction

Nikole D. Patson (Ohio State University) & E. Matthew Husband (University of Oxford)
Huffman.689@osu.edu

A well-established finding in sentence production is that mismatches in the grammatical number of head and intervening noun phrases (NP) lead to agreement errors on the verb (e.g., Bock et al., 2004). This effect is strongest when the head NP is singular and the intervening NP is plural as in: *the key to the cabinets were…* Similarly, although ungrammaticality typically causes disruption in measures of sentence comprehension, the disruption is reduced when the intervening NP has a plural feature. Cue-based memory retrieval accounts propose that the increased number of agreement errors in production and the grammatical illusions in comprehension result from successful retrieval of a partially matching NP (Wagers, Lau, & Phillips 2009). An alternative "noisy channel" approach (e.g., Gibson et al., 2013) predicts that in general the more plural features present around agreement formation the more likely comprehenders may erroneously come to believe that the head NP was plural, leading them ultimately to misinterpret the number of the head NP. We tested this alternative with a question-answer paradigm that has been used to probe misinterpretations (Christianson et al., 2001).

Experiment 1 examined whether comprehenders do indeed misinterpret the number of the head noun when plural features are marked on intervening NPs and verbs. 72 participants read sentences that contained a complex subject NP with a singular head noun. The experiment had a 2 (intervening NP: plural vs. singular) X 2 (verb: plural vs. singular) design (see items 1-4 in Table 1). After reading each sentence, participants were asked a question that probed their final numerical representation of the head noun phrase: *Was there more than one key?* In the experimental items the correct answer to this question was always "no". Table 1 shows the proportion of "yes" responses. As predicted by a noisy channel model, the presence of one plural feature on either the intervening NP or verb lead to significant misinterpretation, and having a plural feature on both the intervening NP and verb lead to significantly more misinterpretation overall.

	Expt. 1	Expt. 2
1. The key to the cabinet is…	.06	
2. The key to the cabinet are…	.22	
3. The key to the cabinets is…	.19	.27
4. The key to the cabinets are…	.35	.47
5. The keys to the cabinets is…		.84
6. The keys to the cabinets are…		.91

Table 1. Proportion of "yes" responses to the question: "Was there more than one key?"

Experiment 2 examined this effect in light of cases where the head noun was marked plural. 52 participants read sentences that contained a complex subject with a plural intervening NP. The experiment had a 2 (head NP: plural vs. singular) X 2 (verb: plural vs. singular) design (items 3-6 in Table 1). As in Experiment 1, when the head noun was singular participants were still likely to misinterpret it as plural when a plural feature was present on either the intervening NP or the verb. Interestingly, even when the head noun was plural, a singular verb interfered with the interpretation, suggesting a symmetric effect of misinterpretation due to mismatching number features.

These results show that when a plural feature is present around agreement formation, comprehenders misinterpret the head NP as plural as evidenced by their increased rate of misinterpretation. This effect is stronger when both the intervening NP and the verb have plural features, consistent with a "noisy channel" account of sentence processing. These results suggest that the increased number of agreement errors in production and the illusion of grammaticality in comprehension may result from comprehenders misinterpreting the head NP as grammatically plural.

Shifting viewpoints and discourse economy

Jesse A. Harris (Pomona College)
Jesse.Harris@pomona.edu

Language users are sensitive to viewpoints other than their own (e.g., Ferguson & Breheny, 2012). Yet, in English, there is no uniform, determinate way to convey perspective, though many terms encode viewpoint, e.g., *beautiful* or *nearby*, described here as the "perspectival center." Rather, a presumptive pragmatic default favoring the speaker's perspective interacts with presentational, surface cues to shift to a non-speaker center (Smith, 2003; Harris, 2012). Such cues include evaluative terms (Kaiser et al 2013), verbs of saying (Harris & Potts, 2009), and subtle prosodic and non-verbal modulations (Harris & Potts, 2011), among others (Smith, 2009). In 3 experiments, I investigate how *syntactic* cues promote perspective shift, finding evidence for the discourse economy principles (1) and (2).

The SPC and NSP discourage perspective shifting generally. However, the discourse processor may use abduction to better interpret an utterance with what is presumptively expected about the speaker and other discourse agents (Hobbs, 1990). Assuming that parenthetical reports (3b) conventionally signal a non-speaker viewpoint (Reinhart, 1983), I predict a preference to maintain the shifted perspective over short stretches of discourse when the tense form maintains a contextual "now", as in the present tense (Klein, 1994).

Experiment 1 tested the effect of parenthetical reports (PRs; 3b) over standard embedding reports (SRs; 3a) in a two-person coordination paradigm, with 15 pairs of subjects (Producer/Listener). After silently judging items like (3) for perspective, Producers verbally performed the item to the Listener. Present tense after PRs elicited more shifted perspectives (73%), while other conditions were at chance. Although Producers were less likely to make a shifted perspective judgment (67%) than Listeners (80%), there was good agreement overall in that participants agreed most on PR-Present items (84%) compared to other conditions (54%).

Experiment 2 tested readers' online expectations for Present or Past following previous reports, PR or SR (N=48). Interpretation questions revealed a greater facilitation for shifted interpretations of the second clause in Present after a PR (d=38%), compared to SRs (d=13%), in addition to independent effects of PR and Present. Reading times showed a crossed interaction: Present was facilitated when following a PR, but penalized after a SR, suggesting that the discourse processor expected a shifted "now", in keeping with the shifted viewpoint. Once the perspective had been shifted, it was economical to maintain, in keeping with NSP.

Experiment 3, a fill-in-the-blank task, compared tense-ambiguous PRs and SRs using modals like *might* (Abusch, 1997; Condoravdi, 2001) in items like (4) with the auxiliary removed from the second sentence. Subjects (N=24) were more likely to provide Present tense auxiliaries (*have, has*) when following PRs (41%) than SRs (24%), despite a general bias against the Present (34%), suggesting that the tense and report type, not the report type alone, promotes a shift.

Conclusion. The form of a report is not sufficient to promote extended perspective shift in discourse: all contextual values must converge on a reading where contextual values are compatible with the shifted viewpoint in order to overturn the SPC. However, once shifted, a shifted perspective may be the most economical, as predicted by NSP.

(1) **Speaker as perspectival center (SPC):** Take the speaker as the perspectival center.
(2) **No shift principle (NSP):** Don't shift perspectives unless required, e.g., by an indicator of incompatible viewpoint.
(3) a. Mary said that there (was | is) a storm. | b. There (was | is) a storm, said Mary. Clouds (had | have) completely covered the sky.
(4) a. Mary said that there might be a storm. | b. There might be a storm, said Mary. Clouds _____ completely covered the sky.

Gestures modulate access to referent representations

Stephani Foraker (SUNY Buffalo State)
forakesm@buffalostate.edu

Recent research indicates that gestured information can influence a comprehender's interpretation of an ambiguous pronoun (Foraker, 2013; Goodrich Smith & Hudson Kam, 2013; see also Cassell, McNeill, & McCullough, 1999). In the present experiments, we investigated whether such off-line findings also held for on-line pronoun resolution, examining reaction times (RTs) for two-alternative forced choice judgments (2AFC), and referent probe recognition. Since the semantic content of gestures is integrated quickly for interpreting action verbs (Kelly, Ozyurek, & Maris, 2010; Özyurek, Willems, Kita, & Hagoort, 2007), we predicted that the gestured information should similarly affect pronoun interpretation during resolution.

Comprehenders viewed 24 short videotaped discourses, which were of a speaker narrating to the camera, using scripted text and gestures. As in the example below, the first part included a gesture accompanying each of two same-gender characters. In the critical last sentence during the pronoun, the gesture was (1) Referent 1 gesture repeated, (2) Referent 2 gesture repeated, (3) Ambiguous 2-handed symmetrical, or (4) No gesture.

Video: "Craig and Matt went on vacation. Craig$_{[\text{Ref 1 gesture}]}$ took a trip to Hawaii while Matt$_{[\text{Ref 2 gesture}]}$ took a trip to Florida. He$_{[1/2/3/4]}$ thought the weather was great while on vacation."

Referent 1 and 2 gestures were performed by different hands, counterbalanced across items. (Both representational gestures with hand shape information or points with location information were included, but no effects of gesture type were found.) A norming experiment established that with only audio input comprehenders showed a strong Referent 2 bias for offline pronoun interpretations. But critically, properties of the speech across the four conditions did not shift interpretation from that biased baseline.

In the 2AFC experiment, Referent 1 and 2 names appeared to the two sides of the video (consistent with gesture direction) at pronoun offset, and the 44 participants chose which person they though the last sentence was talking about. Baseline conditions showed the established recency bias (Referent 2 choice: No gesture 73%, Ambiguous gesture 77%), and a further significant increase with Referent 2 gestures (88%), and a significant decrease with Referent 1 gestures (48%). Furthermore, for Referent 2 name choices, RTs were significantly faster with the consistent Referent 2 gesture compared to the other gesture conditions, ps < .02. Similarly, for Referent 1 name choices, RTs were significantly faster for the consistent Referent 1 gesture vs. the other gesture conditions, ps < .05.

In the referent recognition experiment, a single name probe appeared at pronoun offset, centered above the video, and 32 participants decided whether they recognized the name from earlier in that discourse. An equal number of Referent 1, Referent 2, same-gender foil and opposite-gender foil probes were counterbalanced across lists. Accuracy was consistently high across all conditions. RTs revealed that access to Referent 2 (1593 ms) probes was uniformly faster than for Referent 1 (1922) probes, but also that there was no effect of gesture condition for Referent 2 probes (1518-1700 ms, ps > .49). In contrast, RTs for accessing Referent 1 did differ, with the baseline No (1832) and Ambiguous gesture (1940) conditions in between the faster Referent 1 times (1740) and slower Referent 2 times (2178), linear effect, p = .035.

Taken together, these two experiments show that gestured information can influence online pronoun resolution, shifting the interpretation of an ambiguous pronoun. This evidence is consistent with a multiple constraints approach to anaphor and co-reference resolution (e.g., Kaiser & Trueswell, 2008; Kaiser, 2011), broadening the factors contributing to the cognitive prominence of a referent. Furthermore, the second experiment indicates that gestured information may be a stronger factor in pronoun resolution when the referent representation is more difficult to access in memory (Foraker & McElree, 2007).

Do the pitch accents or the phrasal accents determine an alternative question?

Samantha Heidenreich (Northwestern University), Shari Speer & Carl Polard (The Ohio State University)
samanthagett2015@u.northwestern.edu

Disjunctive questions can be interpreted as yes/no questions or alternative (alt.) questions. The interpretation depends on the intonational contour of the utterance. In “Is Mary allergic to dairy or soy”, if there is emphasis on both disjuncts with a falling pitch at the end of the utterance, the interpretation is (1a); without particular emphasis on the disjuncts, and with a rising intonation at the end of the utterance, the interpretation is (1b).

(1a) Which one of these is Mary allergic to: dairy or soy?

(1b) Is Mary allergic to any of these things: dairy or soy?

These interpretational differences are quite robust (Beck and Kim, 2006), but it has yet to be determined precisely how to characterize the intonational contours themselves. Pruitt and Roelofsen (2011) designed a perception experiment which determined that both the pitch accents (Han and Romero 2004) and final fall (Bartels 1999) of the alternative question are important, but the latter is more important for disambiguation. However, all stimuli in this experiment had the disjuncts at the end of the utterance, thus creating confounding the source of the final fall between a final phrasal accent (L-) and phrasal boundary (L%). In addition, stimuli were laboratory speech, produced with an intentional fixed intonation contour for each of the two interpretations. The stimuli also contained spliced elements, which have been shown to be perceived less accurately than natural stimuli.

Our production and perception experiments were designed to refine this work. Three participants silently read a context paragraph biased toward either a yes/no or alt. question meaning, and then produced a following question or statement as a response with an appropriate intonation. Importantly, the participants never heard the experimenter produce a ‘canonical’ intonation. Each participant produced 24 yes/no utterances and 24 alt. questions. Items were randomized and order of production was counterbalanced. Productions were ToBI annotated. While participants had variable production contours for yes/no utterances (Bartels 1999), the alt. question utterances were more consistent. For 87%, the first disjunct ended high, either in pitch accent (H* or L*+H) or phrasal accent (H-). The pitch accent of the second disjunct was followed by a L- (spreading backward to it if the disjunct was not utterance-final). This supports previous research (Bartels 1999) that classified alt. questions as a coherent class of utterances with respect to utterance final intonational patterns, in contrast to yes/no questions, which may rise or fall at the end.

The productions of one participant were used in a perception experiment. 57 participants read a context paragraph, listened to an utterance, and chose whether the utterance fit naturally with the paragraph. In cases where the intonation of the utterance fit the context, participants correctly judged that it did so 81% of the time for yes/no questions and 91% for alt. questions. Performance was poorer in the inconsistent cases, where participants correctly identified the inconsistency 57% of the time for yes/no questions following alt. contexts and 68% for alt. questions following yes/no contexts. We suggest interlocutors make an effort to fit an utterance into the context. That is, when contexts were appropriate for alt. questions, inconsistent yes/no questions were more easily accommodated. We suggest that it is easier to accommodate a yes/no question, which has a variety of contours, into an inconsistent context than it is to accommodate an alt. question, which has a specific contour, into an inconsistent context.

Bartels, Christine.1999. *The intonation of English statements and questions: a compositional interpretation.* Routledge.
Beck, Sigrid, and Shin-Sook Kim. 2006. Intervention effects in alternative questions. *The Journal of Comparative Germanic Linguistics* 9:165-208.
Han, Chung-hye, and Maribel Romero. 2004. Disjunction, focus, and scope. *Linguistic Inquiry* 35: 179-217.
Pruitt, Kathryn and Floris Roelofsen. 2013. The interpretation of prosody in disjunctive questions. *Linguistic Inquiry* (in press).

Predicting L2 lexical pitch accent perception: The role of top-down knowledge in L2 Japanese

Seth Goss (The Ohio State University) & Katsuo Tamaoka (Nagoya University)
goss.33@osu.edu

Research on L2 lexical perception in pitch/tone-accented languages has suggested that novice listeners rely on bottom-up learning mechanisms, such as acoustic sensitivity and verbal working memory (WM), to aid speech acquisition [1]. L1 listeners, however, use lexical-based (i.e., top-down) knowledge to perceive speech [2]. For advanced L2 listeners, however, it is unclear where they stand on this phonetic-lexical continuum. The current study examined advanced L2 Japanese listeners' ability to identify and categorize 3- and 4-mora words by their lexical accent pattern. We predicted that at the advanced level, listeners no longer rely on basic cognitive resources and instead utilize multiple top-down strategies to aid perception such as L2 lexical knowledge, accent pattern frequency and L1 tone experience.

Sixty advanced L2 Japanese listeners from two L1s, 30 Chinese (+lexical tone) and 30 Korean (-tone), were matched for proficiency and length of study/residence in Japan. Predictor variables were divided into basic cognitive abilities (acoustic pitch sensitivity and verbal WM) and top-down resources (L2 lexical knowledge and +/-tone in the L1). Accent pattern frequency was also used in a post-hoc analysis based on the corpus analysis in [3]. For the lexical accent perception tests, 48 Japanese nouns (12 frequency-matched words from each of 4 possible accent patterns), followed by a postposition and embedded in a short sentence (e.g., *hachiji-ni okiru* 'I wake up at 8') were used to construct two tasks. 1) Identifying if the Noun+Pp was spoken with the correct accent pattern (PitchID) and 2) For the correctly accented items only (n=24), categorizing the Noun+Pp using visual pitch contours in a 4AFC task (PitchCAT).

We found that the L1 Chinese group was significantly more accurate on the PitchID task (78%) than L1 Korean (59%) ($t=7.744$, $p<.001$). On the 4-choice PitchCAT task, Chinese (60%) again showed higher accuracy than Koreans (49%) ($t=2.392$, $p<.05$). Stepwise multiple regression analyses showed that L1 ($\beta=.74$, $p<.001$) and L2 lexical knowledge ($\beta=.37$, $p<.001$) predicted 64% of the variance in PitchID; for PitchCAT, L2 lexical knowledge ($\beta=.36$, $p<.01$) and L1 ($\beta=.32$, $p<.01$) predicted 22% of the variance. By contrast, acoustic sensitivity and verbal WM were not significant contributors to perception performance. Post-hoc analysis of test items revealed accuracy orders on the PitchID task in L1 Chinese to exactly match the frequency order found in [3], i.e., Pattern 4 > P1 > P2 > P3 and were similar in L1 Korean (P4 > P2 > P1 >P3). For PitchCAT, Chinese accuracy again matched [3], but not Korean (P1 > P2 > P3 > P4).

Our regression models showed that advanced L2 listeners exploit multiple sources of top-down knowledge when perceiving lexical pitch accent, including experience with lexical tone in the L1, general L2 lexical knowledge and accent pattern frequency. Contrary to bottom-up perception models [1], which indicate L2 listeners' reliance on basic cognitive resources, the current results suggest this assumption no longer holds in advanced listeners.

[1] Wong, P. & Perrachione, T. (2007). Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics,* 28, 565-585. **[2]** Bent, T., Bradlow, A. R., & Wright, B.A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 97–103. **[3]** Kitahara, M. (2001). Category structure and function of pitch accent in Tokyo Japanese. PhD dissertation. IU-Bloomington.

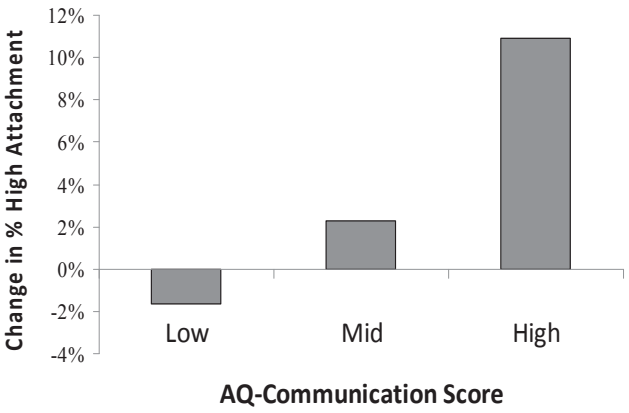
Priming of Implicit Prosody and Individual Differences

Sun-Ah Jun (UCLA) & Jason Bishop (CUNY Staten Island and The Graduate Center)
jun@humnet.ucla.edu

In a sentence such as *Someone shot the servant of the actress who was on the balcony*, it is ambiguous whether the relative clause (RC) modifies NP1 *the servant* (i.e., high attachment) or NP2 *the actress* (low attachment). Although the details of attachment preference are language-specific (Fodor 1998, Fernández 2003), it is known that, cross-linguistically, attachment decisions are sensitive to the sentence's prosodic characteristics, including the location of a prosodic boundary. This fact has been used to support the Implicit Prosody Hypothesis (IPH; Fodor 1998, 2002), which holds that the human sentence parser favors low attachment when the RC forms a single prosodic phrase with NP2, but favors high attachment when a prosodic break directly precedes the RC. Using the structural priming paradigm, the present study explores this hypothesis further by testing whether an implicit prosodic boundary generated from a previously read sentence influences attachment preference for a novel sentence.

In the present experiment, 120 native English speakers reported their attachment interpretations for ambiguous target sentences after silently reading prime sentences that belonged to one of the two conditions, differing in the presence or absence of a comma before the RC. The prime sentence with a comma (i.e., non-restrictive RC condition) would be very likely to be read with a big *implicit* prosodic boundary before the RC (Nespor & Vogel 1986, Watson & Gibson 2004), while this is less likely for the one without a comma (i.e., restrictive RC condition). In both prime type conditions, only a single NP preceded the RC, so the primes themselves contained no attachment ambiguity (and thus no potential for syntactic priming). Based on the implicit prosodic difference between prime type conditions, it was predicted that participants would be more likely to interpret targets as having high attachment if they followed primes in the non-restrictive RC condition. Because previous research has shown that individual differences in "autistic" traits can influence sentence processing (Xiang et al. 2013), and particularly the use of prosody (Jun & Bishop 2013, Bishop 2013), scores on the Autism Spectrum Quotient (AQ; Baron-Cohen et al. 2001) were also collected from participants.

Results of mixed effects logistic regression modeling showed that, as predicted, participants chose high attachment significantly more often after reading sentences in the non-restrictive RC condition (with comma) than in the restrictive RC condition ($p=.035$), thus supporting the IPH. A second model of mixed effects logistic regression which included the participants' scores on each of the subscales of the AQ showed a significant interaction between the prime type and AQ-Communication scores ($p=.021$). Surprisingly, sensitivity to the primes was stronger for individuals with higher AQ-Comm. scores (i.e., those having poorer communication skills); The figure shows the effect of prime type on change in % High attachment as a function of AQ-Comm. We believe this is not because individuals with higher AQ-Comm scores are better at incorporating prosodic information in interpreting the primes, but rather because the presence of a comma interrupted processing, making NP1 more salient in the discourse as the head of the complex NP, thus attaching high. The present results demonstrate that implicit prosodic structure, like syntactic structure, can be primed, and that such priming is in part predictable based on systematic individual differences in the neurotypical population.



Sentence completion to the beat: Effects of implicit prosodic rhythm in English and German

Mara Breen (Mount Holyoke College) & Gerrit Kentner (Goethe University Frankfurt)
mbreen@mtholyoke.edu

Dilley & McAuley (2008) and Niebuhr (2009) demonstrated that English and German listeners interpret stress-ambiguous words/phrases according to a rhythm of alternating strong and weak syllables. Research on implicit prosody (Fodor 2002) suggests that, even in silent reading, the avoidance of *stress clash* (i.e., two adjacent stressed syllables) can guide readers' parsing decisions (Kentner 2012). Here, we report two sentence completion experiments which demonstrate, in English and German, that readers consider sentence rhythm when making syntactic/semantic choices, avoiding both stress clash and *stress lapse* (i.e., two or more adjacent unstressed syllables).

Experiment I: English readers (n=33) provided written continuations of ambiguous sentence fragments (1). Fragments contained a two-word ambiguous region which could be an ADJ-NOUN or NOUN-VERB sequence. The first word was either a monosyllabic (*old)* or disyllabic trochaic (*aging*) word; the stress pattern of the second word varied, with strong-weak stress as a noun (*CONtract*) but weak-strong stress as a verb (*conTRACT*). If stress clash/lapse avoidance is a feature of silent reading, syllable number should influence readers' interpretation of the ambiguity; readers should provide more noun-consistent continuations for (1a) than (1b) and more verb-consistent continuations for (1b) than (1a). Continuations were coded as noun-consistent (n=518), verb-consistent (n=124), or other (n=78). We modeled the proportion of noun responses using a logistic mixed-effects regression with syllable number and the proportion of noun uses of the noun/verb homograph in the Celex corpus as fixed effects, and subjects and items as random effects. The analysis revealed an effect of syllable number, $z = 2.69$, $p<.01$, demonstrating that readers' continuations reflected the established rhythm.

(1)	NOUN-CONSISTENT	VERB-CONSISTENT
a) disyllabic: Nurses think the aging contract…	285 (ideal)	37 (lapse)
b) monosyllabic: Nurses think the old contract…	233 (clash)	87 (ideal)

Experiment II: German speakers (n=161) read gap sentences (2) involving stress-ambiguous verbs like *umSTELLen/UMstellen* (to surround /to shift something) in which the material preceding the gap established a trochaic beat. Participants chose among three options for the gap, which determined the meaning and, crucially, the stress pattern of the verb. The three choices lead to (2a) stress clash, (2b) stress lapse, or (2c) ideal sequences conforming to the established beat. The two options for rhythmically ideal completions (2c) were counterbalanced across lists. For each participant, we computed the probability of choosing the ideal response and excluded two who gave exactly 1/3 ideal responses. A sign test revealed that the number choosing a plurality of ideal responses (n=93) was greater than chance, $p<.05$.

(2) Alex Kohn und seine Leute wollten schnell _____ umstellen.
Alex Kohn and his crew wanted quickly _____ surround / shift

a) clash: den Schrank (UMstellen); engl.: *(shift) the cupboard*	642
b) lapse: die Burgen (umSTELLen); engl.: (*surround) the castles*	682
c) ideal: die Burg (umSTELLen); engl.: (*surround) the castle* / die Schränke (UMstellen); engl.: *(shift) the cupboards*	815

Both experiments confirm the hypothesis that readers consider rhythmic context when choosing among syntactic/semantic alternatives in sentence completion.Together with findings on auditory language processing and reading, these results reinforce the idea that linguistic rhythm is a predictive force in sentence processing, irrespective of modality.

From single words to full sentences

Zeshu Shao & Agnieszka E. Konopka (Max Planck Institute for Psycholinguistics)
Zeshu.Shao@mpi.nl

Sentence formulation unfolds incrementally: people normally begin speaking having encoded some, but not all, of the information they need to produce a full sentence before speech onset. However, the type of information that controls the timecourse of encoding is a matter of debate (Gleitman et al., 2007, vs. Griffin & Bock, 2000). For example, on one account (radical incrementality), speakers may begin formulation by encoding only the first character in an event with priority (e.g., *The fireman...*) and then add information about the second character (...*saving the boy*) after speech onset. On another account (broad-scope planning), formulation may begin instead with speakers encoding information about the event as a whole, which involves encoding information about the action (*saving*). Radical incrementality predicts that formulation should be controlled by the ease of encoding individual characters (i.e., information expressed with *nouns*) while broad-scope planning predicts that the ease of encoding relational information (i.e., information expressed with *verbs*) should also influence formulation.

Here we use a correlational approach to distinguish between these accounts. We test whether production speed and the timecourse of formulation for full sentences can be predicted by speakers' retrieval speed for *nouns* and *verbs*. A stronger influence of noun retrieval speed on sentence formulation would support radically incremental accounts of formulation, while a stronger influence of verb retrieval speed would support broad-scope planning.

22 eye-tracked native speakers of Dutch performed three production tasks in one session. In the first task, they described a series of unrelated pictures, including 33 pictures of transitive, two-character events. In the second and third task, they named 138 object and 103 action pictures, respectively (task order was counterbalanced). These tasks included object and action pictures that elicited the modal nouns and verbs used to refer to the characters and to the actions shown in target events in the event description task, intermixed among filler pictures. The object and action pictures were not visually similar to the target events. Analyses tested whether naming latencies for target nouns and verbs predicted sentence onset latencies as well as eye movements during production of active sentences in the event description task.

Hierarchical multiple regression analyses showed that action naming speed (β = .53, p = .03) but not object naming speed (β = .08, p = .73) predicted sentence onsets (the R^2 of the full model [.32] dropped reliably when action naming was removed [R^2 change = .21, F = 5.78], but not when object naming was removed from the model [R^2 change = .00, F = .12]). Effects of action and object naming speed on the timecourse of formulation were assessed with by-participant quasi-logistic regressions performed over agent-directed fixations (Barr, 2008). Action and object naming did not predict the distribution of fixations in an early, 0-400 ms time window (compared to a simple model including only Time as a fixed effect). Fixations to the agent increased between 400 ms and 1000 ms, and then decreased between 1000 and 1800 ms (approx. speech onset) as speakers began shifting their gaze to the patient. Importantly, shifts of gaze away from the agent were predicted by action naming latencies (producing an Action naming speed x Time interaction), but not object naming latencies.

The results of both analyses suggest that sentence onset latencies as well as the timecourse of formulation are influenced more strongly by processes responsible for encoding relational information than information about individual characters. This supports production accounts of broad-scope planning at the outset of formulation.

References:
Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *JML, 59,* 457-474.
Gleitman, L., et al. (2007). On the give and take between event apprehension and utterance formulation. *JML, 57*, 544-569.
Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psych.Science, 11,* 274–279.

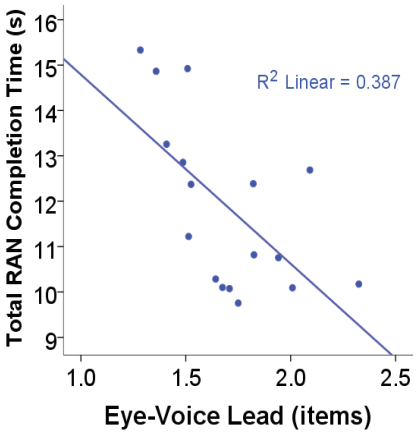
Coordination of eye movements and speech in the RAN task predicts sentence reading

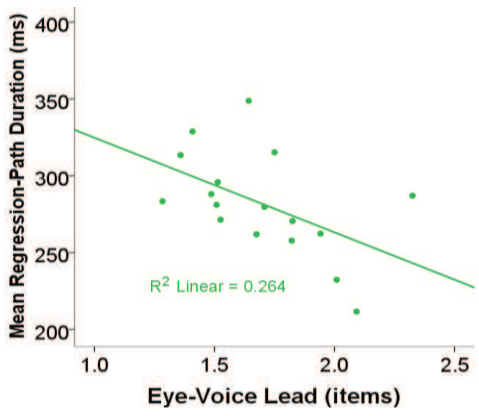
Peter C. Gordon & Renske S. Hoedemaker (University of North Carolina at Chapel Hill)
pcg@email.unc.edu

Efforts to understand individual differences in reading skill have relied heavily on the Rapid Automatized Naming task (RAN), in which the subject is presented with a table of 36 familiar stimuli and must name them out loud as quickly and accurately as possible. Performance is a consistent predictor of reading ability in both children and adults[1,2]. The RAN task requires sustained attention in order to closely coordinate eye-movements, perceptual encoding, working memory, lexical processing and vocal execution, with many of these demands also shared by reading. Practically, optimal RAN performance requires that perceptual-motor processes (eye-movement control, visual attention) be coordinated with somewhat slower processes of articulatory planning and execution so as to avoid overloading working memory. However, the use of total trial completion times, as is customary in the use of RAN as a diagnostic tool, is not informative of how this coordination is achieved on an item-by-item scale. Direct assessment of eye movements and concurrent vocalization reveals how individual differences in eye-mind coordination contribute to efficient processing of written sentences.

Performance on the RAN task was examined in relation to data from the same subjects on a separate task measuring eye movements during sentence reading. Total RAN completion times did not show a robust relationship with eye movements during reading. However, more fine-grained of analysis eye-voice coordination patterns revealed that the extent to which subjects' eyes looked ahead of their voices (ranging from an average of 1.28 to 2.33 across subjects) was an important indicator of processing efficiency on both tasks. Greater look-ahead or eye-voice lead (EVL) was associated with faster overall trial completion times during the RAN, r = -.62, $p < .01$, n = 17, as well as shorter regression-path durations during sentence reading, r = -.51, $p < .05$, n = 17. In addition, effects of word frequency on regression-path duration were attenuated for participants with more efficient EVL, r = -.62, $p < .01$, n = 17. Perhaps surprisingly, EVL was not related to first-pass measures of sentence reading.

Unlike previous studies of the RAN and eye-tracking[3], the present operationalization of eye-voice lead is not affected by vocal duration, providing a measure of eye-voice coordination that isolates attentional control. Our findings suggest that more efficient RAN performance was achieved by those participants with sufficient attentional resources to cope with the higher working memory load resulting from looking relatively further ahead from the current vocalization. During reading, these participants showed more consistent forward progress throughout the sentence and attenuated sensitivity to potential disruption by difficult, low-frequency words,. Taken together, these results show that analysis of eye-voice relations in RAN and similar tasks offers a method for assessing observable differences in attentional control processes that contribute to effective coordination of multiple processes during skilled sentence processing.





[1] National Early Literacy Panel (2008). [2] Kuperman & Van Dyke (2011), *JML*. [3] Jones et al. (2008) *Cognition.*

Poster #1040

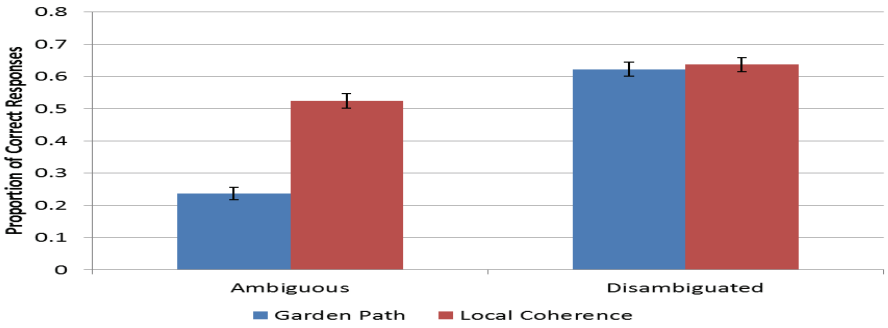
Integrating new vs. revising previous input: Local coherences vs. garden paths

Kiel Christianson (UIUC), Steven G. Luke (BYU), & Kacey Wochna (Ithaca College)
E-mail: <kiel@illinois.edu>

Garden-path sentences (GP; 1a) and local coherence sentences (LC; 1c) both create parsing difficulty, but for different reasons. Here we characterize garden-paths as "early-disruption" structures, in which the reader cannot know that the preferred parse is incorrect until later material is encountered. In contrast, local coherences are "late-disruption" structures, where the parse of the early part of the sentence is ultimately correct, but later material must be adjusted to fit onto the existing, correct partial parse. We explore this theoretically interesting distinction using eye-tracking to directly compare the two structures, comparing both online eye movement data and offline comprehension data. ***Predictions***: Reanalysis is a process of last resort (Sturt et al., 2001), and the parser is loath to relinquish previous interpretations (Christianson et al., 2001, 2006; Ferreira & Henderson, 1991). Therefore, GPs should yield signs of disruption in the eye movement record that are slower to manifest compared to disruptions caused by LCs. Furthermore, initial misinterpretations derived from GPs should be more resistant to revision, and thus lead to higher levels of misinterpretation than in LCs, whose misinterpretations may quickly revised to better match the existing, correct, partial interpretation. ***Methods***: 20 pairs of GP and LC sentences in either ambiguous or unambiguous conditions were created with equivalent semantic content (1a-d). Materials were distributed in a 2x2 Latin square design along with 72 fillers across four lists. 48 participants read the sentences while their eyes were tracked with an EyeLink 1000 eye tracker. After each sentence, participants answered a comprehension question (1e). Verbs were chosen so that lengths and frequencies would work against predictions.

(1) a. The player tossed the ball *interfered with* the other team. (GP, amb.)
b. The player who was tossed the ball *interfered with* the other team. (GP, unamb.)
c. The other team interfered with the player *tossed* the ball. (LC, amb.)
d. The other team interfered with the player who was *tossed* the ball. (LC, unamb.)
e. Did the player toss the ball?

Results & conclusions: Residualized reading times on the italicized critical verbs in (1) were analyzed using repeated measures 2x2 ANOVAs. Robust main effects of ambiguity (ambig > unambig) were observed for first fixation duration, gaze duration, go-past time, and total time. Importantly, significant interactions were also obtained, whereby inflated fixation durations were observed for ambiguous LCs compared to ambiguous GPs in all measures except total time, where the two types did not differ. Logit mixed-models were used to analyze comprehension accuracy. As shown in the figure below, ambiguous LCs were comprehended more accurately than ambiguous GPs. As predicted, LCs triggered faster reanalysis attempts and proved easier to recover from; however, very low comprehension rates for both structures suggest that researchers should be cautious about drawing conclusions about parsing mechanisms without assessing interpretation.



Skipping syntactically illegal "the"-previews - the role of predictability

Matthew Abbott (University of California, San Diego), Bernhard Angele (Bournemouth University), Danbi Ahn & Keith Rayner (University of California, San Diego)
mabbott@ucsd.edu

Readers tend to skip words, particularly when they are short and frequent (O'Regan, 1979; Drieghe, Pollatsek, Staub, & Rayner, 2008), or predictable from the prior context (Balota, Pollatsek, & Rayner, 1985). Angele and Rayner (2013) reported a gaze-contingent boundary experiment (Rayner, 1975) where readers were presented with either a preview of a frequent 3-letter word (*fix*), a nonword (*fde*), or *the*, in sentences where *the* was syntactically infelicitous (You should not *fix* it unless it is broken). Readers skipped *the* when presented parafoveally more often than the target words (first-pass skipping probability, p-skip, .51 vs. .29). Processing difficulty was apparent following *the* previews with increased go-past times on the posttarget word, as well as more regressions out of that region. The results indicated that readers often cannot detect syntactic anomalies in parafoveal vision, although readers did sometimes skip the pretarget word in the *the* preview condition. We replicated this study and manipulated target word predictability to assess whether contextual constraint modulates *the*-skipping behavior.

Readers were presented with 36 sentences in which a 3-letter target word was either predictable from context (e.g., My watch broke and without the right tool I couldn't *fix* it and make it work) or unpredictable (e.g., I couldn't adequately *fix* my watch and make it work without the right tool). Cloze probabilities in the predictable context averaged .75 and .10 in the unpredictable context. An invisible boundary was placed in the space preceding the target word; readers were presented with either a valid preview of the target word (*fix*) or a preview of *the*.

Readers skipped the target word more often in the predictable context compared to the unpredictable context when given a valid preview (p-skip .55 vs. .39). Readers skipped a preview of *the* more often than a valid preview of the target (p-skip .73 vs. .47), and this difference was greater in the predictable than in the unpredictable condition (p-skip .78 vs. .67). When readers fixated the target word, they did not show standard preview benefit effects; first fixations were actually numerically shorter following *the* previews than valid previews (220ms vs. 227ms). Gaze duration showed the same pattern of results. Go-past times on the posttarget word were longer following *the* previews than valid previews (430ms vs. 350ms), and were not modulated by predictability. Readers regressed more often out of the posttarget word following *the* previews than valid previews (.33 vs. .18).

The present study provides further evidence that readers frequently skip the article *the* when infelicitous in context. Readers skipped predictable words more often than unpredictable words, even when *the*, which was syntactically illegal and unpredictable from the prior context, was presented as a parafoveal preview. Interestingly, readers displayed sensitivity to predictability regardless of the preview, as they skipped *the* more often in the predictable condition. However, readers suffered similar consequences when skipping *the*, as reflected by a general inflation of reading times and an increase in regressions out of the posttarget word in both predictability conditions. Thus, it appears that when a short word is predictable in context, a decision to skip it can be made even if the information available parafoveally conflicts both visually and syntactically with those predictions. Further research will be conducted to determine whether these results are exclusive to the word *the*, or extend to other parafoveal previews.

Angele, B., & Rayner, K. (2013). Processing *the* in the parafovea: Are articles skipped automatically? *JEP: LMC*, *39*(2), 649-662.

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17,* 364–390.

Drieghe, D., Pollatsek, A., Staub, A., & Rayner, K. (2008). The word grouping hypothesis and eye movements during reading. *JEP: Learning, Memory, and Cognition, 34,* 1552– 1560.

O'Regan, J. K. (1979). Eye guidance in reading: Evidence for the linguistic control hypothesis. *Perception & Psychophysics, 25,* 501–509.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cog. Psych., 7,* 65–81.

Predicting individual differences in underspecification: An integrated model of good-enough processing

Felix Engelmann & Shravan Vasishth (University of Potsdam)
felix.engelmann@uni-potsdam.de

The theory of good-enough processing (Ferreira et al., 2002) suggests that readers strategically adapt their parsing efforts to task demands and available resources, with the consequence of omitting processes that are less important either for the current goal (Swets et al., 2008) or for building a coherent structure (as postulated by Construal, Carreiras & Clifton, 1993). Indeed, von der Malsburg & Vasishth (2012) have shown that readers with low working memory capacity leave an ambiguous adjunct attachment underspecified more often than high-capacity readers. It is not clear, however, how this adaptation works. Is it necessary to assume that low-capacity readers use a different parsing strategy, or can the difference be explained by a common mechanism? So far there is no detailed model of the good-enough account that could clarify this issue.

We propose that underspecification results from a common strategy that aims for an uninterrupted reading process and therefore constrains attachment effort, when possible, by the timing of eye movement control processes. We implemented a model where the process of determining the attachment site of an ambiguous adjunct is subject to this interaction between parsing and eye movements. Specifically, the attachment process gets canceled if it has not reached its final stage by the time the next word has been identified and is ready for integration. A varying behavior due to working memory differences would be predicted by the assumption that low-capacity readers on average take longer to complete the attachment than high-capacity readers, resulting in more cancelations that leave the attachment underspecified. Eye movements were simulated in ACT-R using the eye movement module EMMA (Salvucci, 2001) interacting with a cue-based retrieval parser (Lewis & Vasishth, 2005). Individual differences in working memory capacity were modeled in accordance with previous studies (e.g., Daily et al., 2011) by randomly varying ACT-R goal buffer source activation W (normally distributed with mean 1 and SD 0.25; higher values of W improve speed and accuracy of retrieval processes). Low capacity was defined as $W \leq 1$ and high capacity as $W > 1$. Following Swets et al. (2008), an underspecification is not corrected later in the sentence. If an attachment was made, however, contradicting disambiguation information leads to a repair operation triggering a regression towards the beginning of the sentence.

We tested the model predictions against the results of von der Malsburg & Vasishth (2012). In their eye-tracking study of temporarily ambiguous (high vs. low) adjunct attachment in Spanish, they found (i) an ambiguity advantage in gaze durations in the pre-disambiguation region, driven mainly by low-capacity readers, and (ii) an increased probability for high-capacity readers to reread the sentence after disambiguation in the dispreferred high-attachment condition. They concluded that high-capacity readers attach more often and, hence, make more wrong attachments that require a reanalysis.

We simulated 50 participants reading the three experimental conditions (high, low, unambiguous) 15 times each. EMMA parameters were set to values estimated in Engelmann et al. (TopiCS, 2013) using the Potsdam Sentence Corpus (Kliegl et al., 2004). Other ACT-R parameters were set to values used previously by Lewis & Vasishth (2005).

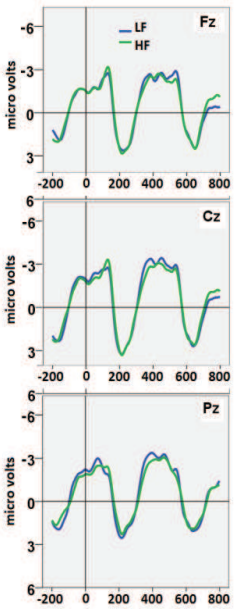
The simulation results showed that the proposed interaction of parsing, eye movement control, and working memory correctly predicts (i) the observed ambiguity advantage in gaze durations and (ii) the difference in rereading proportions in the high-attachment condition. The presented simulations, thus, support the good-enough account in ambiguous adjunct attachment and demonstrate that a single mechanism can lead to different rates of underspecification modulated by working memory capacity.

Recognizing Words during Sentence Processing: Behavioral and Neural Measures

Peter C. Gordon, Giulia C. Pancani, Matthew W. Lowder, Renske S. Hoedemaker & Mariah Moore (University of North Carolina at Chapel Hill). pcg@unc.edu

The process of recognizing words during reading is influenced both by the characteristics of the words to be recognized and by the contextual information provided by the sentences in which the words appear. Results from eye-tracking during reading have shown additive effects between sentence context and word frequency, while these same factors have yielded a variety of patterns in ERP studies. The current study addressed this difference by measuring the effects of word frequency on eye movements and ERPs using the same subjects and stimuli. Subjects participated in two sentence-reading tasks, one in which eye movements were recorded during normal reading and a second in which EEGs were recorded as sentences were presented one word at a time using RSVP (ISI of 400 ms, 300 ms on and 100 ms off). As shown, the frequency of target words (5 to 9 letters) at two sentence locations (early and late) was manipulated such that the frequency of the high-frequency member of a pair was at least one unit of log frequency greater than that of the low-frequency member of the pair. Two hundred sentences were constructed so that the sentence frames provided minimal constraint on the identity of the target words; average length of the high and low frequency target words was the same, and across subjects they appeared in the same target locations. 40 sentences were used in the eye-tracking task yielding 40 observations for each subject in each frequency condition (20 at each location). The remaining 160 sentences were used in the RSVP ERP task, yielding 160 observations for each subject in each frequency condition (80 at each location). Assignment of sentences to task was varied across the 16 subjects so that all sentence frames and target words contributed equally to the eye-tracking and ERP results. *The majestic* (HF: *painting*; LF: *tapestry*) *was displayed in the* (HF: *lobby*; LF: *foyer*) *of the* …

For the eye-tracking study, frequency of the target word had highly reliable facilitative effects on the first-pass eye-movement measures that are generally taken to indicate word recognition: The rate of first-pass skipping was greater for high-frequency than low-frequency targets (.168 *vs* .109; $F = 7.3$, $p < .025$); gaze duration (the sum of the durations of all first-pass fixations) was shorter for high-frequency than low-frequency targets (247 *vs* 270 ms; $F = 20.1$, $p < .001$); finally, word frequency did not interact with target location in the sentence. Unlike the strong effects found for first-pass reading measures, ERPs were only marginally affected by word frequency in the RSVP sentence processing task (see figure). No statistically significant differences between the conditions were found for the N250 time window (200-300 ms, $F < 1$), which is thought to reflect ortho-phonological encoding. The N400 time window (300-500), which is thought to reflect retrieval of lexical-semantic information, showed a slight trend toward a smaller N400 for high-frequency words ($F = 3.66$, $p = .076$). Follow-up analyses examining a number of different epochs and electrodes also failed to show other differences between conditions or any interactions between word frequency and target location.



Frequency effects were very robust in first-pass eye-movement measures of word recognition but were marginal in the ERP study despite its having four times as many observations per condition and subject as the eye-tracking study, a difference that is not attributable to stimuli or subjects. This substantial difference in sensitivity to word frequency during sentence processing adds to well-established differences in the timecourse of language processing as measured by eye movements and ERPs. Reconciling these divergent results is necessary in formulating an integrated understanding of the behavioral and neural characteristics of sentence processing.

Selective priority for structure in memory retrieval

Dan Parker & Colin Phillips (University of Maryland)
dparker3@umd.edu

The role of structural information in guiding retrieval operations in sentence comprehension is controversial. Evidence from multiple tasks and phenomena has been used to argue that structural information has a privileged status in retrieving the heads of dependencies [1,2,3]. However, widespread evidence from non-structural interference effects suggests that structural priority is more limited. This contrast has motivated various proposals regarding the memory architecture for the parser, including the use of qualitatively different access mechanisms, or the selective use of structural cues for different dependencies. **1.** We show that for anaphor resolution, a phenomenon that has resisted non-structural interference effects in the past, we can systematically control where interference effects do and do not occur. **2.** Furthermore, we show based on computational simulations that it is possible to derive both the presence and absence of the interference effect from the same retrieval mechanism. **3.** We propose an account of why it is harder to obtain interference effects in anaphor resolution than in some other dependencies, based on whether retrieval is triggered by error correction or normal resolution, and we provide supporting timing evidence from three eye-tracking studies.

1. Selective anaphor illusions: Facilitatory interference effects manifest as eased processing of illicit words due to the retrieval of a structurally irrelevant licensor, leading to illusory licensing. Previous studies on reflexive anaphors have failed to consistently find such effects, but they tested only contexts in which the illicit dependency involved one feature mismatch, e.g. gender [1] or number [3]. Across three eye-tracking experiments we compared contexts involving a 1-feature mismatch, e.g., gender or number, and a 2-feature mismatch, e.g., gender+number (1a), gender+animacy (1b), number+animacy (1c). Additional conditions involving agreement attraction [as in 6] provided a baseline measure of interference. Results show that reflexives are indeed susceptible to interference, but only selectively: when the reflexive mismatches the true subject in just one feature, there is no interference, but when it mismatches in two features, strong facilitatory interference is found, comparable in size to agreement attraction.

(1) Expt a [n=30]. The {librarian | janitor} said that the schoolboys reminded **herself** …
Expt b [n=30]. The {librarian | janitor} said that the memo reminded **herself** …
Expt c [n=30]. The memo that the schoolgirl(s) noticed reminded **themselves** …

2. Computational simulations: ACT-R simulations [4] suggest that the contrasting interference effects for reflexives may reflect a weighted combinatorics scheme in which structural cues are weighted more strongly in retrieval than morphological cues. Crucially, our simulations revealed weights that can capture both the presence and absence of interference from the same model. However, this raises the question of why agreement shows strong interference even in a 1-feature mismatch context, implying that structure is not weighted more strongly for agreement.

3. Deriving variability across dependencies: Why should structural cues be prioritized for anaphora but not for agreement? We suggest that whereas structural cues are prioritized in retrieval in the normal course of processing unpredictable dependencies (e.g., anaphora, thematic binding [2]), they are not prioritized when retrieval is triggered by an unfulfilled prediction (e.g., agreement violations), leading the parser to be uncertain about the accuracy of its existing structural encoding. In support of this account, the eye-tracking data shows different timing profiles for interference, which is observed early (first-pass) for anaphora, but only as a late effect (total time) for agreement.

[1] Sturt. 2003. *JML.* [2] Van Dyke & McElree. 2011. *JML.* [3] Dillon et al. 2013. *JML* [4] Lewis & Vasishth. 2005. *Cog. Sci.* [5] Trommershauser et al. 2011. OUP. [6] Wagers et al. 2009. *JML.*

Aim low: Speakers design utterances for the most naive addressee

Si On Yoon & Sarah Brown-Schmidt (University of Illinois, Urbana-Champaign)
syoon10@illinois.edu

In multiparty conversation, the degree of common ground (CG) between dyads within a larger group often differs. Previous studies of dyads show that speakers use longer referential expressions to accommodate the needs of a naive listener (Wilkes-Gibbs & Clark, 1992). How these processes scale up to multiparty conversation, where dyads within the larger group differ in joint knowledge, is unknown. Here we test several hypotheses about how speakers design utterances in such cases: (1) Speakers may estimate the average knowledge state of all addressees and design expressions with respect to the average. (2) Speakers may design expressions for the person with whom they have the most common ground, at the expense of naïve addressees, or (3) for the person with the least common ground, at the expense of efficient communication with knowledgeable addressees. (4) When alternating between one of two addressees, speakers can flexibly draw on distinct representations of common ground.

We explored these questions in triadic conversation. Three naïve participants (N=60) were randomly assigned the role Director (D), Matcher 1 (M1), or Matcher 2 (M2). The task consisted of two phases: a sorting task and test. D and M1 first played a sorting game, entraining names for game-pieces as a dyad, followed by testing as either a dyad or a triad. During sorting, M2 performed unrelated tasks in another room. Thus, D&M1 had high CG, and D&M2 had low CG. At test, each person had a separate computer that showed 4 pictures from the sorting on each trial. The key manipulation was who participated in the conversation at test:

Matcher 1: D instructs M1 only, while M2 is in another room. *[High common ground (CG)]*
Matcher 2: D instructs M2 only, while M1 is in another room. *[Low CG]*
Matcher 1&2 together: D simultaneously instructs M1&M2 in same room *[Mixed CG]*
Matcher 1/2 alternating: D alternates instructing M1 / M2, both in same room *[high/low CG]*

In each of the four conditions, D instructed one addressee (Matcher 1, Matcher 2, Matcher 1/2 alternating conditions), or two addressees (Matcher 1&2 together) to click on one of the 4 objects. In the alternating condition, the addressee (M1 or M2) clicked the target; the other matcher simply hit enter to advance the trial. All target items were repeated twice to examine how quickly D adjusted descriptions to the current addressee(s). Conditions were within-subjects; 4 different experimental blocks rotated participants through the four conditions.

RESULTS: We analyzed the length of D's referential expressions using maximal mixed-effects models; results are significant at $p<.05$ unless noted. Replicating previous findings (Brennan & Clark, 1996), Ds produced significantly longer expressions for naïve partners (**M2 condition**) than partners with whom they had formed common ground (**M1 condition**), $t=6.0$. There was no difference between the **M2** and **M1&2 together** conditions: Ds designed long expressions any time M2 was an addressee, showing that Ds do not design expressions based on average addressee knowledge (Hyp#1), nor the highest addressee knowledge (Hyp#2). Instead, Ds design expressions for the most ignorant addressee (Hyp#3). When describing the target a second time, expressions in the **M2** and **M1&2 together** conditions were significantly shorter ($t=4.92$), showing rapid formation of common ground. Finally, consistent with Hyp#4, Ds flexibly designed expressions to the knowledge of the current addressee in the **M1/2 alternating condition**, using shorter descriptions for M1 than M2 ($t=6.07$).

CONCLUSION: Every-day conversation often involves multiple parties with different knowledge, yet mechanisms of language processing in multiparty conversation are largely unexplored. We show that speakers flexibly alternate between partner-specific representations of common ground, and when addressing multiple parties, speakers tailor descriptions for the least knowledgeable person. To achieve such flexibility, speakers likely maintain and use distinct representations of common ground for multiple parties. Such sophistication is inconsistent with egocentric-default theories of speaking (e.g., Horton & Keysar, 1996).

Definiteness and givenness affect pronoun resolution: Evidence from eye fixations

Klaus von Heusinger, Sofiana Chiriacescu, Andreas Brocher, & Tim Graf (Universität zu Köln)
sofiana.chiriacescu@uni-koeln.de

Current models on anaphora resolution have shown that various phonological, syntactic, and semantic properties of an antecedent noun phrase (NP) can affect the interpretation of a subsequent pronoun (Arnold, 2001; Kehler et al., 2008). However, no study has investigated to what extend (i) *definiteness* (i.e., definite vs. indefinite NP) or (ii) *givenness* (i.e., brand-new vs. linked reference) affect a referent's availability for subsequent re-mention. We provide empirical evidence that both factors increase the activation status of a referent, and should thus be included in models on anaphora resolution.

It is generally assumed that a definite NP *presupposes* its descriptive content and the uniqueness of its referent, while an indefinite NP *asserts* the existence of its descriptive content and makes a non-uniqueness implicature (Heim, 2011). Furthermore, definite and indefinite NPs introduce discourse referents that have different (cognitive) information statuses generally associated with different degrees of givenness: *given*, *linked* (also '*bridged*' for definites) to the discourse, or *brand-new* (Prince, 1981; Strube & Hahn, 1999). Given NPs refer back to an explicitly introduced referent. Linked NPs require an *inference* and a certain level of relatedness between their anchor and the anaphor (Table 1 (2 a-b): a gym class context activates a script including the referents 'instructor' and 'participant'). In contrast, brand-new NPs do not rely on an inference relation (Table 1 (2 c-d)). Investigating the likelihood of pronominalisation, previous research has underlined the preferential status of given information, while brand-new and linked referents have been widely neglected.

Table1: Sample experimental materials (translated from German)

(1) The gym class was crowded as usual.	(2) Michael observed_____	(3) When the door was closed, **he** put his towel on the mattress and started to stretch.
	(a) the committed instructor. [def+linked]	
	(b) a committed participant. [indef+linked]	
	(c) the American near the loudspeaker. [def+brand-new]	
	(d) an American near the loudspeaker. [indef+brand-new]	

We investigated whether the definiteness and givenness of a NP that introduces a discourse referent for the first time affect the interpretation of a pronoun as being co-referent with this NP. We predicted that both factors should increase the activation status of a referent, such that a subsequent pronoun is resolved to it. In a visual world eye-tracking experiment participants (*n=24*) listened to short stories (Table 1, (1) – (3)), while their eye fixations were monitored. All critical referents appeared in direct object (DO) position and were manipulated with respect to their morphological realization and information status, yielding four conditions: (i) definite and linked (2a); (ii) indefinite and linked (2b); (iii) definite and brand-new (2c); (iv) indefinite and brand-new (2d). Association strength to contexts was normed and was similar for referents in (2a) and (2b), *t*s < 0.3, *p*s > .7. The last sentence contained an ambiguous pronoun, *he,* which could refer back to the subject or the DO of the preceding sentence.

Results of separate linear mixed-effects models (0-1200ms; 0-400ms; 400-800ms; 800-1200ms) analysing proportions of looks to the DO over the subject revealed that, within 1200ms post pronoun onset, participants were more likely to interpret the pronoun as the preceding critical referent when that referent was definite rather than indefinite, $t = 2.53$, $p = .011$. In addition, within the first 400ms, a main effect of *Definiteness* was accompanied by a marginal *Definiteness* X *Givenness* interaction, $t = -1.93$, $p = .053$, indicating that DO referents were initially more strongly fixated when they were definite *and* linked. We conclude that: (i) definiteness is a reliable predictor of how likely it is that a pronoun is interpreted as co-referring with a preceding NP and that (ii) givenness shows an early effect on pronoun resolution only for definite NPs. The results provide novel evidence that various semantic-pragmatic dimensions, such as givenness and definiteness, interact *differently* with definite and indefinite NPs.

Differences in the prosodic encoding of information structure in closely-related languages

Arunima Choudhury and Elsi Kaiser (University of Southern California)
arunimac@usc.edu

We investigated how speakers use prosody to encode information structure in two flexible word-order languages, Hindi and Bangla. Proponents of interactive designs (e.g. Ito/Speer'06, Brown-Schmidt et al'08) emphasize the importance of collaborative processes in conversation. Building on these insights, we used an interactive task to investigate prosodic marking of information structure in two related Indo-Aryan languages, Hindi and Bangla. Previous work on English has led to divergent claims about the mapping between focus types and f0 contours: Some claim that different pitch contours map onto different focus types (e.g. Pierrehumbert'80), but others argue that the mapping between prosody and focus types is not one-to-one (e.g. Watson et al.'08). We aim to shed more light on the mapping between focus types and prosodic cues from a cross-linguistic perspective. Specifically, we investigated the role of fundamental frequency (f0) in encoding New-Information focus, Corrective focus and Selective focus (illustrated below):

(1) A: What color did Peter paint his bicycle? B: He painted it [blue]. NEW INFO FOCUS
(2) A: Peter painted his bicycle red. B: No, he painted it [blue]. CORRECTIVE FOCUS
(3) A: Did Peter paint his bicycle red or blue? B: He painted it [blue]. SELECTIVE FOCUS

We compared **Hindi and Bangla**, to see how/whether they use f0 to mark focus types. Both Hindi and Bangla are SOV languages with relatively free word order. They mark focus (i) positionally at the default focus position (immediate preverbal position); (ii) morphologically using focus particles; and (iii) prosodically. Although prior work has shown that the absence/presence of focus in Hindi and Bangla can be signaled prosodically (Hayes & Lahiri 1991, Fery 2010, Khan 2008, Patil et al 2008), little is known about *whether different focus types have different prosodic realizations*. This issue is important for the debate about whether focus is a semantically unified concept or has distinct subtypes (e.g., Rooth'92, Schwarzschild'98, Katz/ Selkirk'11, Hartmann/Zimmermann'06). If a language distinguishes the focus types prosodically, this would seem to suggest that they should be treated as distinct subtypes.

Our research has 3 main aims: **First**, we explored whether Hindi and Bangla speakers use prosodic cues (specifically, fundamental frequency f0) in a similar manner to distinguish focused vs. unfocused constituents. **Second**, we tested whether they distinguish the three focus types (ex.1-3) prosodically. **Third**, we tested both canonical (SOV) and non-canonical (OSV) orders to see how syntax and prosody interact. Two studies were conducted with native Hindi (n=25) and Bangla speakers (n=17). We manipulated focus type (New-Information, Corrective, Selective), word order (SOV/OSV) and grammatical role of the focused element (Subject, Object) for a 3x2x2 design. (Selective focus=selecting from a mentioned set, e.g. Zimmermann/Onea'11). We used an interactive task to maximize naturalness; participants worked in pairs, narrating stories to each other. Targets were embedded in the stories.

Results show that both Hindi and Bangla use f0 to differentiate between *focused vs. unfocused subjects and objects*, but only at the default preverbal focus position. Both languages also use f0 to indicate post-focal compression. However, w.r.t focus type, Hindi shows no reliable f0 differences *between focus types* irrespective of word order or grammatical role (p's=n.s). In contrast, in Bangla, the f0's for *different focus types* do differ significantly for both subjects and objects in SOV and OSV, but only at the default focus position (p's<.05; Corrective >Selective>New-Information). **In sum**, both languages convey presence/absence of focus with f0, but only at the default focus position. *However, whereas Bangla uses of f0 to make fine-grained distinctions between focus types, Hindi does not.* Thus, although the studies were parallel and on related languages, there are differences in the mapping between focus types and prosody. This highlights crosslinguistic differences in information structure encoding, and provides new data for a tripartite distinction between new-info, corrective and selective focus.

Dissociating neural effects of predictability and incongruity in adjective-noun phrases

Ellen Lau & Anna Namyst (University of Maryland, College Park)
ellenlau@umd.edu

While much work has demonstrated evidence that comprehenders use the context to generate predictions about upcoming input, an unresolved debate is the extent to which the processing mechanisms that are facilitated by lexical prediction are the same as those that are involved in resolution of semantic and pragmatic anomaly. In particular, the amplitude of the N400 ERP response has been argued to be modulated by both lexical-semantic predictability and semantic/pragmatic incongruity. However, these factors have been difficult to fully disentangle because predictability and semantic/pragmatic well-fittedness are often correlated, and the constraints of the off-line completion tasks used to estimate predictability make it challenging to precisely control predictability between relatively less predictive sentence contexts. Here, in two ERP experiments, we dissociated these factors by using adjective-noun sequences, for which the probability of the noun given the adjective can be estimated by using bigram counts from large corpora. In the first experiment we held well-fittedness constant and modulated probability, and in the second we held probability constant and modulated well-fittedness. We observe effects of both predictability and incongruity in the N400 time-window, but with qualitatively different spatial distributions.

In Experiment 1 (n=36), all adjective-noun sequences were semantically and pragmatically well-formed, and we contrasted the response to nouns for which the probability was low or high given the previous adjective, e.g. *runny nose* vs. *dainty nose*. In Experiment 2 (n=32), all nouns had a low probability ($p < .005$) and all adjectives had low constraint, and we contrasted the response to nouns that either resulted in a semantically/pragmatically congruent or incongruent phrase, e.g. *raw apple* or *intelligent citizen* vs. *intelligent apple* or *raw citizen*. Adjective-noun onset asynchrony was set at 600ms, and a memory recognition post-test ensured attention.

Qualitatively different ERP effects were observed for the predictability and the well-formedness contrasts. In Experiment 1, low probability nouns elicited a more negative response than high probability nouns between 300-500ms over central-posterior electrodes, resulting in a significant interaction between anterior-posterior distribution and predictability ($p < .05$). In Experiment 2, incongruent nouns also elicited a more negative response than congruent nouns between 300-500ms ($p < .05$), but the response was broadly distributed with a slightly anterior focus, such that there was no interaction between congruity and anterior-posterior distribution ($p > .2$). In the subsequent 600-800ms time-window the distribution of sustained negativity was similarly reversed, with increased posterior negativity for less predictable nouns in Experiment 1, and increased anterior negativity for incongruent nouns in Experiment 2, resulting in significant interactions between anterior-posterior distribution and predictability/congruity in both ($ps < .05$).

These results suggest that the resolution of semantic/pragmatic incongruity involves processes that are at least partially dissociable from those processes that are modulated by lexical-semantic predictability. We speculate that the anterior negativity observed to incongruent nouns may be related to semantic reanalysis either of the kind used to process figurative language (Coulson & Kutas, 2001) or semantic coercion (Pylkkänen et al. 2009).

Coulson, S. & Kutas, M. (2001). Getting it: Human event-related brain response to jokes in good and poor comprehenders. *Neuroscience letters, 316*(2), 71-74.
Pylkkänen, L., Martin, A., McElree, B., & Smart, A. (2008). The anterior midline field: Coercion or decision-making? *Brain and Language, 108*(3),184-190.

Effects of context on the processing of adversative and comparative constructions

Grégoire Winterstein (Aix-Marseille Université, Nanyang Technological University), Emilia Ellsiepen (Paris 7, Universität Frankfurt), Jacques Jayez (ENS Lyon) & Barbara Hemforth (CNRS)
gregoire.winterstein@linguist.univ-paris-diderot.fr

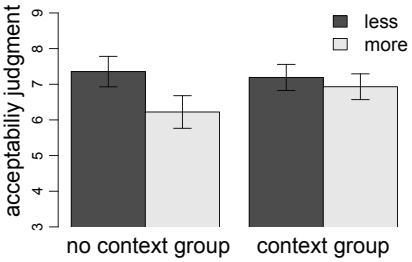
This work focuses on the combination of adversative connectives (e.g. *but*) with comparative constructions. The basic observation is that, when introduced by an adversative, a comparative expression can only indicate inferiority; mentioning superiority leads to infelicity (1). Replacing *but* by *and* makes the contrast vanish, confirming that *but* is partly responsible for it.

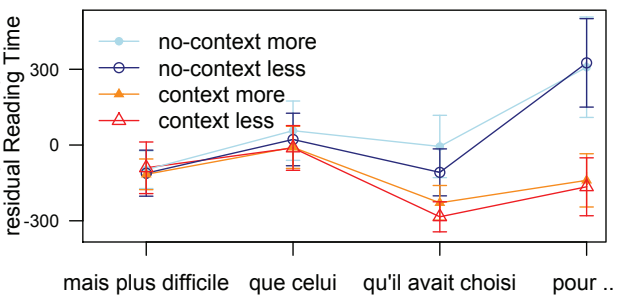
1. a. The Friday exam was difficult, but less difficult than the Tuesday exam.
   b. # The Friday exam was difficult, but more difficult than the Tuesday exam.

Intuitively, the contrast is less sharp if those examples are used in a facilitating context. For (1b) such a context could be that a teacher has two classes and is supposed to give them equally difficult exams. Asking whether the teacher achieved this balance makes (1b) sound more acceptable as an answer by making explicit the opposition necessary for interpreting *but*.

Theories of *but* deal with this picture in different ways. *Relevance Theory* (RT) [1] considers that a *but*-conjunction requires an element that is *accessible* from the first conjunct and world-knowledge and that gets *denied* by the second. Without context, it is assumed that this element cannot be found in (1b). In context, the use of *but* is perfectly felicitous since this element is explicitly made accessible. In *Bayesian Argumentation Theory* (BAT) [2], the contrast in (1) is accounted for on semantic grounds. The infelicity of (1b) is explained by considering that both conjuncts are lower-bounding expressions and thus that their probabilistic effects are similar: their assertions modify degrees of belief in comparable ways, hampering the abduction of the opposition necessary for *but*. Therefore, even with a context, (1b) is expected to differ from (1a) since the semantic argumentative clash remains present, even if context overrides it.

In order to test the differences between RT and BAT, we conducted a region-by-region self-paced reading experiment immediately followed by an acceptability judgment study in French. The twelve experimental items consisted of a test sentence and a context designed to rise the expectation of some property being equal for two entities (as in (1)). Test sentences were of the same form as in (1), in the *more* condition the comparison was formed with *mais plus + adjective*, while in the *less* condition it was formed with *mais moins + adjective*. Half of the participants, the *no-context* group, were presented the test sentence in isolation, while the other half, the *context* group, saw the test sentence preceded by the facilitating context. After completing the reading part, participants were asked to judge the sentences' acceptability on a 9 point scale. For the *context* group, sentences were again preceded by the context. Analysis of the acceptability ratings using mixed models indicated both, a main effect of condition ($\chi(1) = 8.14, p < .01$), and an interaction of condition and group ($\chi(1) = 4.59, p < .05$). A follow-up analysis showed that the effect of condition was only significant in the *no-context* group, confirming the remedial effect of context on the acceptability of sentences like (1b). For the self-paced reading data, we analyzed residual reading times from the critical region containing the comparison, as well as the three following regions to allow the context to take effect on reading times. Condition had a significant effect in region 3 ($\chi(1) = 4.00, p < .05$), which followed the comparison object, where reading times were longer in the *more* condition. Here, the interaction with group was not significant. This suggests that while context affect offline measures, online difficulties arise both, with and without context, as predicted by BAT.





References: [1] Blakemore D. (2002) *Relevance and Linguistic Meaning. The Semantics and Pragmatics of Discourse Markers,* Cambridge U. Press, Cambridge.
[2] Winterstein G. (2012) "What *but*-sentences argue for: a modern argumentative analysis of *but*", *Lingua* 122 (15).

106

Elaboration of concepts facilitates their retrieval in sentence processing

Melissa Troyer (UCSD), Philip Hofmeister (University of Essex), and Marta Kutas (UCSD)
mtroyer@ucsd.edu

Language comprehension involves the retrieval of ir
ation of that information in the service of comprehension. Current models of sentence processing claim that relevant cues at retrieval sites enable access to appropriate targets in memory [1, 2]. Further, models of similarity-based interference predict that to the extent that a target has more unique features cued at retrieval, retrieval will be facilitated (due to less interference) [3]. Such models are consistent with recent findings that retrieval of an element within a sentence is facilitated by a greater number of syntactic and semantic cues. Indeed, reading more complex noun phrases, such as *the Texas cattle rancher*, as opposed to less complex noun phrases, such as *the rancher*, lead to faster reading times at later points within the same sentence [4]. However, no study has investigated the retrieval of concepts that have been more or less elaborated by other types of material in larger discourses. The current study extends previous work by elaborating elements (here, descriptions of people) in short texts. We hypothesized that effects of greater elaboration (leading to more semantic cues) would persist across sentences to facilitate retrieval in a subsequent target sentence. In our experiment, 48 participants read 24 short texts like (1). Sentences 1-2 were presented in their entirety one at a time, and sentence 3 was presented word-by-word using self-paced reading (slashes indicate regions used for analysis). All short texts were followed by a comprehension question.

(1) Two senators were arguing with a Democrat and a Republican after a big debate.
The Democrat had voted for one of the senators, and the Republican had voted for the other, a man from Ohio who was running for president.
The senator / who the { Democrat / Republican } / had voted for / was picking a fight / about health care reform.

Critically, sentence 2 depicted two individuals from sentence 1 (with the same definite noun phrase descriptor, here *senator*) with either many ("many-cue" condition) or only one ("one-cue" condition) additional piece(s) of information. In sentence 3, the subject of the embedded relative clause picked out which of the two senators was the object of *supported* (Democrat="one-cue", Republican="many-cue").[1] We found that the many-cue condition led to reduced reading times in both the retrieval region of the relative clause (*had voted for;* β = .097, p < .01) and, marginally, in the final region of the sentence (*about health care reform*; β = .038, p = .10), supporting our hypothesis that more elaborate representations lead to facilitated access. In addition, scores on the Author & Magazine Recognition Tests [5], a proxy for reading ability and possibly world knowledge, were inversely correlated with reading times at the retrieval region (r = -0.32, p < .05), but only marginally so at the final region (r = -0.26, p = .08). However, accuracy on questions about the second sentence (containing the elaborative information) was positively correlated with reading times at the final region (r = 0.28, p = .05), but not the retrieval region (r = .06, p = .69). These findings suggest a possible differentiation of information sources (e.g., a general, perhaps verbal, ability and a task-driven ability) which may contribute to the effects of increased elaboration on ease of retrieval. We propose that what and how much one knows about a concept can impact retrieval in sentence processing, as in other basic memory paradigms. Though many models currently consider memory during sentence processing to operate under many of the same principles as general human memory processes [1-4], these models may benefit from incorporating information about richness of concepts and knowledge at a level beyond the scope of single sentences, including variation both from individual concepts and also from individual subject variability in knowledge and verbal expertise.

References: [1] Van Dyke & McElree, 2011; [2] Lewis & Vasishth, 2005; [3] Gordon, Hendrick & Johnson, 2001; [4] Hofmeister, 2011; [5] Stanovich & Cunningham, 1992

[1] Items were constructed with four versions, including two versions reversing the order of the senators in sentence 2, where greater amount of elaboration came first. This manipulation (order) was entered into analyses and produced a single main effect of order at Region 2 (with more recently encountered material being read more quickly; β = .060, p
effect may result from initial ease of processing for more recently encountered entities.

Hippocampal contributions to discourse processing: Findings from Amnesia

Sarah Brown-Schmidt (University of Illinois), Melissa Duff & Jake Kurczek (University of Iowa)
sarahbrownschmidt@gmail.com

The goal of the present research is to examine the contribution of the hippocampal dependent declarative memory system to the establishment and use of discourse representations during on-line processing of reference. Recent findings outside the domain of language show that the hippocampus contributes to processing of relations and maintenance of information even over very brief time-periods (Hannula, et al., 2006; Öztekin, et al, 2010). Here we test the hypothesis that the use of discourse cues to guide on-line reference resolution is hippocampus-dependent. Our approach is to examine the on-line processing of short discourses in individuals with amnesia who have bilateral hippocampal damage and severe declarative memory impairment.

Method: Combining eye-tracking and neuropsychological methods, eye-tracked participants (3 amnesic patients and 7 matched healthy comparison participants) viewed a scene while listening to dialogs that first introduced two same-gender characters as in [1]. In one condition the 1st-mentioned character (e.g., Mickey) was made more prominent through re-mention as in [2]. An unrelated item was then mentioned in [3] followed by the critical pronoun, e.g., "he" (underlined) in [4]. We tested interpretation of the critical pronoun.

[1] Mickey is painting a portrait of Donald,
[2] [Mickey is trying really hard to get the portrait just right, because he wants to be a famous artist someday] [3] and some paint is spilling on the floor.
[4] And what is he wearing? Look, he's wearing yellow shoes.

The design was modeled on previous research (Arnold, et al., 2000; Song & Fisher, 2005), and manipulated (a) whether the 1st mentioned character was repeatedly mentioned as in [2], or not, and (b) whether the pronoun referred to the 1st or 2nd mentioned character. The intended referent of the pronoun was disambiguated given the sentence + scene (e.g., at *yellow shoes*).

Results: Eye-movements during interpretation of the critical pronoun (underlined) were analyzed using maximal mixed models, and included Group (Amnesia vs. Comparison), Repetition (1st repeated vs. not), and Mention (pronoun refers to 1st or 2nd-mentioned character) as factors. Effects are significant at $p<.05$ unless noted. Consistent with previous findings, there was a significantly larger target preference when the pronoun referred to the 1st mentioned character. This effect was amplified with repetition, indicated by a significant Mention*Repetition interaction. These effects were qualified by interactions with Group. Separate analyses by Group revealed that, as expected, comparison participants exhibited significant effects of both Mention, and a Mention*Repetition interaction.

Amnesic participants showed a different pattern, indicating that even in these brief dialogues, developing and using a representation of the discourse drew on declarative memory. While amnesic participants did show a significant effect of Mention, it was smaller in magnitude than that of healthy comparison participants. The fact that they did show a Mention effect is noteworthy because it shows that at least some aspect of the larger discourse representation was intact (e.g., that Mickey is more prominent). Amnesic participants did not, however, show a significant Mention*Repetition interaction, demonstrating that, unlike healthy comparisons, they failed to use this discourse cue to boost salience of the 1st-mentioned referent when that referent was mentioned three times during the preceding discourse.

Conclusion: The pattern of successes and failures during on-line processing in amnesic participants shows that while coarse-grained representations of referent salience were used, more subtle aspects of the discourse history were not integrated into on-line referential processing. These findings offer key support for the hypothesis that language processing depends on contributions of hippocampus. These contributions may be particularly great when flexible integration of multiple information sources is necessary, even over short time periods.

Impact of focus alternatives on memory for focus referents

Wind Cowles & Michelle Perdomo (University of Florida)
cowles@ufl.edu

One influential theoretical approach to focus proposes that focus status indicates the presence of alternatives that are relevant for interpretation (Féry & Krifka, 2008; Rooth, 1992). The interpretation (and thus processing) of focus crucially involves not only the focus referent itself, but also the set of alternatives to that referent. There is evidence suggesting that focus alternative sets are activated as part of focus processing (e.g. Benz et al., 2012; Fraundorf et al., 2010; Ito & Speer, 2008), with focus alternatives showing priming with auditory stimuli (e.g. Husband & Ferreira, 2012) as well as lower detection accuracy in a text-change paradigm (e.g. Cowles et al., 2012).

This study examines the impact of focus alternative sets on memory for a focused referent in a discourse, and in particular whether processing the focus alternatives interferes with memory for a focused referent. Twenty-eight participants each read 120 short passages and after each one confirmed whether a paraphrase of one of the passage's sentences was accurate. In 40 critical trials, the passages were comprised of five carefully constructed sentences that introduced five discourse referents, created an explicit two-member focus alternative set and selected a focus from it. The paraphrase was either accurate (Same), or had one word substituted with another word from the passage that was either the explicit focus alternative (Alternative) or was mentioned but not given as an explicit focus alternative (Mentioned). Accuracy and response times for the paraphrases were recorded.

(1) A soprano and a critic were waiting backstage after a performance. A cellist and a violinist were nearby, talking with a composer about the next show. The critic asked the soprano who she admired more, the cellist or the violinist. The soprano replied that it was the cellist that she liked. The critic agreed he was quite something.
Same Paraphrase: The soprano admired the cellist.
Alternative Paraphrase: The soprano admired the violinist.
Mentioned Paraphrase: The soprano admired the conductor.

If the focus alternative (e.g. *violinist*) is considered during focus processing, it could interfere with readers' ability to subsequently recall the precise identity of the focus referent. Based on this, we predicted that responses to the Alternative Paraphrase would be the longest and least accurate. These predictions were partially borne out, with pair-wise t-tests showing Alternative Paraphrases having significantly ($p < .001$) lower accuracy than Same or Alternative Paraphrases, but showing no difference in response times:

Paraphrase Type	RT (StErr)	Accuracy (StErr)
Same	2869 (126)	91.2% (1.6)
Alternative	2761 (107)	78.7% (3.1)
Mentioned	2692 (118)	91.4% (1.8)

These results are consistent with prior work suggesting that focus alternative sets are part of focus processing, and also show a lingering effect on memory for focus referents, suggesting that while focus referents usually show an advantage with respect to memory, they are not immune to interference from referents that may have been initially considered as part of focus processing. Work is currently underway to establish how long this interference continues, and whether focus alternatives create long-lasting uncertainty about the identity of the focus referent.

Is EVERY child different until they turn 15?: Not so quick development of universal quantifier comprehension

Kiwako Ito (Ohio State University), Utako Minai (University of Kansas) & Adam Royer (Ohio State University)
ito@ling.ohio-state.edu

Young children often fail to accept statements such as “Every pig is carrying a basket” as a correct description of a scene where the number of objects involved in the action (baskets) exceeds the number of agents (pigs). Minai et al. (2012) showed that (1) 4- and 5-year-olds often incorrectly rejected *every*-statements when an extra object was present in the scene, whereas (2) these children fell into two groups: those who consistently rejected and those who consistently accepted these *every*-statements. Importantly, those who incorrectly rejected these *every*-statements showed frequent looks to the extra object, while those who correctly accepted the *every*-statements (both children and adults) rarely fixated on it. The present study examines responses to *every*-statements from a wider age range of English speakers to test (a) whether the division among individuals described in (2) is a characteristic of only young children, and (b) whether the local prosodic prominence can change the response patterns described in (1).

A total of 123 native English speakers (age: 4-69) saw 36 slides in the experiment. In 12 critical slides, three animals carried out the same action with an object while the last quadrant contained an extra object (Fig1). Participants viewed the slide for 2.5 seconds and heard an *every*-sentence describing the scene. All sentences were recorded with three prosodic patterns: with an emphatic pitch accent on either *every* or the agent noun, and without an emphasis on any word. Participants judged whether each sentence correctly described the scene by pressing the YES or NO button. When the NO button was pressed, the screen showed “Why?” to elicit their verbal justification for rejection.

The aforementioned split was revealed not only among preschoolers but also among older participants up to 14-years-old. Those who were 15-years-old and above mostly accepted the *every*-statements. Also, children who rejected the *every*-statements (n=22) fixated on the extra object more frequently than those who accepted the statements (n=53) (Fig2 vs. Fig3: $p<.01$). Interestingly, the effect of pitch prominence was found only in children who incorrectly rejected the statements; when *every* was emphasized, fixations to the extra object were significantly reduced compared to the other two conditions ($p<.01$) during the object NP. Thus, the results suggest that some children may take years before achieving adult-like *every* comprehension. While the prosodic emphasis on *every* can momentarily redirect their attention to the agents or their action, its impact on comprehension seems weak (as they still ended up rejecting the *every*-statements). Once children are able to comprehend *every*, prosody seems to play little role.

The findings may speak to the debate regarding whether children’s comprehension errors can be attributed to (i) their non-adult-like linguistic knowledge about *every* (Philip, 1995), or to (ii) extra-linguistic factors, such as children’s still-developing executive function (Minai et al., 2012). Our findings revealed that older children (up to 15) split into those who can correctly interpret *every* and those who cannot. We take these findings to be most consistent with the extra-linguistic account as in (ii) (Minai et al., 2012), given that the development of executive function has been argued to continue into adolescence (Best et al., 2009).

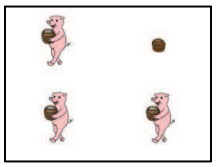


Fig 1: Sample slide for the statement 'Every pig is carrying a basket'

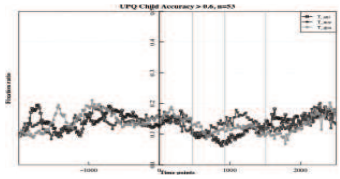


Fig. 2: YES group

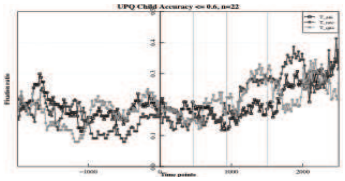


Fig. 3: NO group

Pragmatic processing costs reflect linking to context, not enrichment

Shevaun Lewis (Johns Hopkins University) & Colin Phillips (University of Maryland)
shevaun@jhu.edu

In the real-time comprehension of scalar expressions (SEs) like 'some', implicature-enriched interpretations often take longer than literal interpretations **[1-4]**. This observation seems to support the hypothesis that pragmatic enrichments are generated—at some cost—only after the literal interpretation is rejected. However, several recent studies have failed to find implicature-related processing cost **[5-7]**. We argue that the different results are due to critical differences in the contexts: increased costs arise when the quantified set was not previously mentioned. We hypothesize that the observed costs reflect the process of accommodating the existence of the set and inferring its role in context. Using a paradigm similar to **[2,3,7]**, we demonstrate that the cost of processing SEs is affected by the difficulty of linking to context, not by pragmatic enrichment. This holds not only for 'some', but also for ad-hoc scales. There were also differences between the two scale types, challenging the widespread assumption that findings about 'some' are representative of scalar implicature in general. We conclude that, at least for relatively conventional implicatures, there is no cost for accessing the enriched meaning above what is required for accessing relevant contextual information.

METHODS: Participants' (n=24) eye movements were recorded as they read 3-sentence passages. The final sentence contained the critical SE, which was either 'some' (1) or a member of an *ad-hoc scale* (2). The first sentence established the Context Type by mentioning the *upper bound* (UB) or *lower bound* (LB) of the scale, or neither (*neutral* condition). The middle sentence added distance between the first mention of the scale and the critical SE to avoid potential repeated name effects **[6,7]**. The regions of interest were the Trigger (the SE), Complement ('the rest'), and spillover regions for each. We calculated five measures of reading time (RT)—first pass, right-bound, regression path, reread, and total time—and used linear mixed effects models to analyze the effect of Context on RT separately for each Scale Type.

(1) *Some/all*: Professor Smith needed to {grade all of the papers/grade some of the papers/be productive} tonight despite her exhaustion. She worked for several hours after dinner. At 10:00 she emailed her TA to say that she had graded some of the papers before getting too tired, but the rest would have to wait until tomorrow.

(2) *Ad-hoc*: Professor Smith needed to {grade even the longest papers/grade at least the shortest papers/be productive} tonight … she had graded the shortest papers before getting too tired, but the rest would have to wait until tomorrow.

RESULTS: RT at the Complement is diagnostic of the interpretation of the SE: processing 'the rest' should be facilitated when the SE is upper-bounded **[2,3,6,7]**. With the *some/all* scale, RTs were lower in *UB* contexts at the Complement (regression path, reread time) and its spillover region (total time), suggesting that comprehenders had computed an upper-bounded interpretation of 'some'. With the *ad-hoc* scale, there were no significant differences in these regions. Thus, there was no evidence that interpretations of the *ad-hoc* SE differed based on Context, at least during the time frame of our observation.

If processing costs for SEs were related to enrichment, we would expect RTs at the Trigger to show the opposite pattern from the Complement: for the *some/all* scale, RTs would be higher in *UB* contexts. In fact, RTs in the Trigger region were higher in the *neutral* contexts for both the *some/all* scale (right-bound, regression path, total time) and the *ad-hoc* scale (first pass, rightbound, total time). These results support our hypothesis that processing costs are related to linking the SE to the context—specifically, inferring a relevant set.

[1] Bott & Noveck (2004). *JML*, 51, 437-457. [2] Breheny, Katsos & Williams (2006). *Cognition*, *100*, 434-463. [3] Bergen & Grodner (2012). *J. of Exp Psych.: Learning, Memory, and Cognition*, *38*(5), 1450-1460. [4] Bott, Bailey, & Grodner (2012). *JML*, 51(3), 437-457. [5] Feeney et al. (2004). *Canadian J. of Exp. Psych*., *58*(2), 121-132. [6] Hartshorne & Snedeker (submitted). [7] Politzer-Ahles & Fiorentino (2013). *PLoS ONE*, *8*(5): e63943.

Prosodic encoding of information structure depends on frequency and probability

Iris Chuoying Ouyang & Elsi Kaiser (University of Southern California)
chuoyino@usc.edu

Introduction: Prosody reflects the informativity of linguistic elements. Prior work has approached informativity from two angles**: (i)** information theory/statistical probability and **(ii)** information structure /pragmatics. From the information-theory side, increased statistical probability (e.g. high frequency, contextual predictability) has been found to reduce prosodic prominence (e.g. Aylett/Turk'04, Gahl'08). From the information-structure side, information-structural salience has been found to increase prosodic prominence (e.g. words are more prominent in narrow than wide focus, e.g. Breen et al '10). However, little attention has been paid to the *interaction* between these two dimensions (cf. Baker/Bradlow'09). This was our focus.

Method: We investigated whether **word frequency** and **contextual probability** modulate the prosodic effects of information structure, and whether they impact different categories of information structure in different ways. In a production study, participants (N=16) worked with a lab assistant as their partner: the partner asked questions and participants answered. Three question types were used to elicit focus: **new-information narrow focus** (e.g. *What did Rita and Mary kick in the garage?*), **corrective narrow focus** (e.g. *I heard that Rita and Mary kicked dirt in the garage.*), and **VP (wide) focus** (e.g. *What did Rita and Mary do?*). In the responses (e.g. *They kicked cars in the garage*), target objects differed in **(a)** word frequency (high, e.g. *cars/books*, or low, e.g. *cans/shells*) and **(b)** contextual probability (high, e.g. *cars/cans*, or low, e.g. *books/shells*). Contextual probability was estimated for each item/context in a norming study; word frequencies are from SUBTLEXus. Earlier work suggests low-frequency words show smaller prosodic differences between focused and unfocused information (Baker/Bradlow'09). So, we **predicted** that distinctions between focus types would be masked with infrequent *or* contextually improbable words: We expect *the prosodic effects of information structure to be weakened when other factors also demand prosodic prominence.*

Analysis: We tested 3 regions: pre-focus (verb), focus (object), post-focus (prep+article, e.g. *in the*). We included post-focus regions because pitch lowering right after focused elements occurs with narrow focus (e.g. Eady/Cooper'86). Pitch contours were compared at the alpha=0.05 level with Smoothing Spline ANOVA models. **Results:** Pre-focus and focus regions showed no pitch differences. However, in the *post-focus region*, frequency and probability modulate effects of information structure on f0: **(a)** When the focused word is *high-frequency and contextually highly probable*, new and corrective narrow focus have significantly lower pitch than VP focus. This fits with pitch lowering signaling narrow focus. **(b)** When the focused word is *high-frequency but contextually improbable*, only new-information focus has significantly lower pitch; corrective and VP focus do not differ. **(c)** When the focused word is *low-frequency but contextually highly probable*, only corrective focus has significantly lower pitch; new-information focus patterns with VP focus. **(d)** When the focused word is *low-frequency and contextually improbable*, all three focus types pattern alike. **In sum,** effects of narrow focus **(i)** start to disappear when another factor also demands prosodic prominence, i.e. when focused words are infrequent or contextually improbable, and **(ii)** completely disappear when there are two other sources of prosodic prominence (low frequency *and* low probability). Interestingly, corrective focus is impacted more by contextual probability, and new-information focus by word frequency. (We show this can be derived from speakers' expectations about interlocutors' mental states, suggesting an interaction-based account.)

Our work provides new insights into prosodic prominence by combining statistical (word frequency, contextual probability) and pragmatic factors (information structure). ***As a whole, our results show that the prosodic encoding of information-structural categories depends on the word's frequency and contextual probability.*** Crucially, many existing studies have examined probable contexts and have not manipulated frequency, which may explain why narrow focus has been found to bring greater prosodic prominence than wide focus.

Restrictive vs. non-restrictive composition: An MEG study

Timothy Leffel, Miriam Lauter, Masha Westerlund & Liina Pylkkänen (New York University)
ml187@nyu.edu

Introduction. Research on language processing has shown that the meanings of complex phrases such as modified nouns are constructed incrementally during comprehension. Language users are thus able to integrate the meanings of individual words as they are encountered, resulting in rapid association between phrases and the objects in the world that they refer to. This association is often called "reference resolution." Numerous studies have shown that when speakers encounter a noun modifier, their default assumption is that it is "contrastive;" it narrows the reference set of objects to which the phrase will refer (Sedivy et al., 1999). However, this assumption can be overridden if a "non-contrastive" modifier is supported by the pragmatic context or the specific communicative goal (Sedivy, 2003). For example, in *I visited my Italian friend*, *Italian* is used to pick out a specific individual from a set of friends (restrictive), whereas in *I visited my sick mother*, *sick* would typically not serve to pick out a mother from a set (non-restrictive). In the non-restrictive case, the modifier simply conveys extra information about the scenario. The goal of this study was to determine whether brain responses hypothesized to reflect composition are sensitive to the restrictiveness or non-restrictiveness of composition. Recent neurolinguistic investigation has implicated the left anterior temporal lobe (LATL) as a central region for basic composition. For instance, the LATL shows greater activity for a noun in a combinatory context than the same noun in a non-combinatory context (Bemis & Pylkkänen, 2011). However, because the stimuli of these studies have lacked contextual information, they have not addressed the sensitivity of the LATL to contextual factors. Employing a simple question-answer paradigm, we manipulated whether a prenominal adjective or determiner is interpreted restrictively. We used MEG to compare brain activity between the same phrase in restrictive vs. non-restrictive contexts.

Materials. Questions were followed by a two or three word answers. To achieve a minimal contrast in restrictiveness, we varied question-type while holding the stimulus material of the answer constant (1-2). A parallel strategy was used to manipulate the restrictiveness of prenominal determiners (3-4) (the bolded noun was the target of MEG analysis):

(1) Q: *Which chicken should the farmer slaughter next?* A: *His fat **chicken**.* (restrictive)
(2) Q: *Will the farmer slaughter his chicken or his lamb?* A: *His fat **chicken**.* (non-restrictive)
(3) Q: *Will the farmer slaughter his chicken or Mary's chicken?* A: *His **chicken**.* (restrictive)
(4) Q: *Will the farmer slaughter his chicken or his lamb?* A: *His **chicken**.* (non-restrictive)

Results. The LATL, as well as its right hemisphere homologue, was indeed sensitive to restriction, with restrictive composition eliciting higher responses than non-restrictive composition. However, this effect was only observed when the restricting element was a determiner. Adjectival stimuli showed the opposite pattern, which we hypothesize to be driven by the special pragmatic properties of non-restrictive adjectives (providing an explanation for what is asserted in the rest of the sentence). Overall, our results demonstrate a robust sensitivity of the LATL as well as the right anterior temporal lobe to high level contextual and potentially also pragmatic factors. These findings are inconsistent with a syntactic interpretation of the LATL but do conform to theories treating it as a multi-faceted semantic hub where many kinds of meaning-related tasks are performed.

References

Bemis, D. K., & Pylkkänen, L. (2011). Simple Composition: An MEG investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience*, 31(8).
Sedivy, J. C. (2003). Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of psycholinguistic research*, 32(1), 3-23.
Sedivy, J.C., Tanenhaus, M.K., Chambers, C.G., & Carlson, G.N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, 71, 109-147.

Soft and hard presupposition triggers are fast in online processing

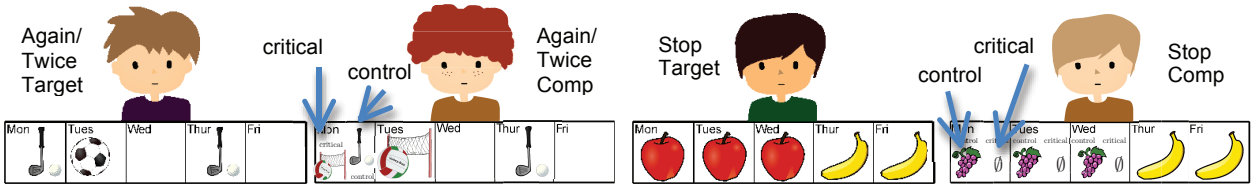
Florian Schwarz (University of Pennsylvania)
florians@ling.upenn.edu

Introduction: To date, only a couple of studies have provided detailed time-course information on presupposition processing, e.g., in eyetracking in reading in Schwarz & Tiemann (2012) and in a visual world studies on *also* by Romoli et al. (2012). We extend these efforts by investigating *again* and *stop*, providing further evidence for rapid presupposition processing. The equivalence of the two in processing is of theoretical relevance given proposals for distinguishing different classes of triggers. A minimal comparison with *twice* (equivalent to *again* but with no presupposition) furthermore compares asserted and presupposed content directly.

Experiment 1: We used a reference identification task with four figures - depicted on the screen with a schedule of activities – one of which had to be chosen as matching a linguistic description (1): Two opposite gender distractors, a target that matched the description throughout, and a competitor that either did or did not meet the condition introduced by *again*/*twice*.

(1) *Some of these children went to play golf on Monday, and some to play volleyball.*
*John went to play golf {**again** later on /**twice** this week} and also played soccer on Tuesday.*

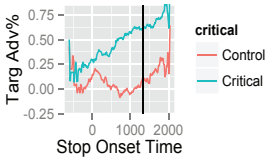
In the critical condition, the target could already be identified during the otherwise ambiguous time window (see underlining), based on *again*/*twice*. The last-mentioned activity always provided independent disambiguation. 27 participants saw 24 such sentence-picture pairs. LMEM analyses of Target Advantage Scores (TAs) for the critical conditions revealed an immediate shift to target from the earliest point on (200ms after *again*/*twice* onset) for both *again* and *twice*, suggesting that the relevant implication was immediately available, and indistinguishably so in the presupposed and asserted conditions.



Experiment 2: The presupposition of *stop* (that the relevant activity had been going on before) was investigated with the same paradigm. Disambiguation during the underlined part of the sentence was again possible in the critical condition based on the presupposition, though final disambiguation occurred independently later on:

(2) **Context**: *These children got nice treats for their snacks this week.*
Target: *Henry stopped eating the delicious apples on Thursday.*

Eye movements exhibited a significant shift in TAs right upon encountering *stop*, indicating that this presupposition is also available right away.



Discussion: The results broaden the evidence for the online processing of presuppositions. They are inconsistent with accounts that predict a delayed availability for (at least certain) presuppositions due to their assumed pragmatic derivation. In the case of *again*, the comparison with *twice* shows directly that the time-course is parallel to asserted content. The results also contribute to recent debates about potential differences between classes of presupposition triggers, e.g., hard (here *again*) vs. soft (here *stop*) ones, (Abusch 2010; see also Cummins & Amaral 2013 a.o.). While these may need to be distinguished for other reasons, their online processing time-course does not provide independent motivation for doing so. Finally, the present approach opens up new methodological avenues for investigating a host of important theoretical questions that require evidence beyond the level of intuitions.

The role of working memory in the online realization of scalar inferences

Stephen Politzer-Ahles (New York University Abu Dhabi), Robert Fiorentino (University of Kansas), Jeffrey Durbin (University of Kansas), & Lillian Li (University of Kansas)
spa268@nyu.edu

Statements like “some of my relatives are coming”, are often interpreted as “not all of my relatives are coming”. The nature of the process that derives this sort of interpretation is under debate: some accounts propose that the derivation of this reading is automatic and effortless, whereas others propose that it is a pragmatic inference that requires additional cognitive resources over and above realizing the basic semantic interpretation. Sentence judgment experiments have shown that working memory, as well as processing load imposed by a concurrent task, influence how comprehenders ultimately interpret scalar quantifiers (Marty & Chemla, 2013, *inter alia*). These data, however, cannot easily distinguish between processing costs incurred by making the upper-bounded interpretation and processing costs incurred by evaluating the meaning of that interpretation with respect to the explicit task. It is valuable, therefore, to investigate the nature of these processes online and with more implicit measures.

We measured the relationship between individual-level cognitive abilities and the interpretation of scalar expressions in a self-paced reading experiment (based on Breheny et al., 2006) that tests whether readers have implicitly realized an upper-bounded interpretation, without explicitly asking for their interpretation. Participants read vignettes in which the context asks a question about *all* of something (which biases readers towards assigning an upper-bounded reading to “some”) or about *any* (which biases them against such a reading):

1) Mary asked John whether (*all/any*) of his relatives were staying in his apartment. John said that <u>some of them</u> were. He added that <u>the rest</u> would be staying in a hotel.

Previous research has shown that participants read “the rest” more quickly in a context where they have assigned an upper-bounded interpretation to “some” (because the interpretation “not all of the relatives are staying in his apartment” makes the reader aware of an additional subset of relatives referred to by “the rest”). In the present study, 48 volunteers completed this self-paced reading task, as well as a battery of cognitive measures including two working memory span measures (reading span and count span), Flanker and Stroop tasks, the Autism Spectrum Quotient, and the Interpersonal Reactivity Index. We tested whether these individual differences predict the extent to which readers realize upper-bounded interpretations of scalar terms.

Composite working memory was a significant predictor ($p < .005$) of how the sentences were interpreted. The typical effect of inferencing in reading times at “the rest” (faster reading times in the “all” context than the “any” context) was observed in participants with higher working memory, indicating that these participants assigned upper-bounded interpretations in the former context more than the latter; on the other hand, participants with lower working memory did not show this effect. This interaction was not observed in control items where the quantifier was “only some” instead of “some” (i.e., items where the upper-bounded interpretation is semantically specified and thus should not require any additional pragmatic inferencing to be realized). These results suggest that realizing upper-bounded interpretations of scalar expressions may be an effortful process that requires working memory resources, and they extend the findings of offline studies by providing evidence that the locus of these processing costs is indeed in the inference-making process itself, rather than in the later evaluation of this meaning with respect to a task.

Word order typology and discourse expectation

Eunkyung Yi & Jean-Pierre Koenig (University at Buffalo)
eyi@buffalo.edu

The most common way to mark a *causal* relation is to use markers such as *because, since,* etc. Dryer (1992) has shown that such markers tend to occur clause-initially in verb-initial (VO) languages and clause-finally in verb-final (OV) languages. Diessel (2001) has additionally shown that causal clauses with a clause-initial marker tend to be flexible in their position (they can precede or follow the *result* clause), whereas causal clauses with a clause-final marker predominantly precedes the *result* clause. As a consequence, the 'non-iconic' order of presentation of a causal relation (the *result* clause precedes the *causal* clause, counter to the order of events in the world) is more common in VO languages than in OV languages. The present study examines the impact of such word order correlations on speakers' discourse expectations in sentence continuations.

Our hypothesis is that even when no explicit clause markers are present, speakers of VO languages like English tend to use more often the non-iconic causal order than speakers of OV languages like Korean because speakers are influenced by the grammars of their language and the relative frequency of iconic and non-iconic causal orders. We further expect that the typological differences between the two languages will affect narrative discourse, but not conversational discourse. This is because causal markers and relative order of cause and result clauses play no role across speech turns and will therefore not impact speakers' discourse expectations. To test our hypothesis, we constructed English and Korean sentence prompts that were equivalent in meaning. Each language prompt was presented in two different discourse types, narrative and conversation, as illustrated in (1).

(1) Examples of English stimuli (translation-equivalent sentences to Korean speakers)
   a. Narrative discourse
      Gloria hated Bill. _____
   b. Conversational discourse
      Jessica: Gloria hated Bill.
      You: _____

A distinct group of participants was assigned to each language/discourse-type condition (two groups of participants in each language were assigned to two discourse types, respectively). To guard against null effects, we manipulated verb meaning within each group: Verbs were either implicit causality verbs (IC) or non-Implicit causality verbs (non-IC), as IC verbs are known to significantly increase non-iconic result-cause order in English (Kehler, 2002). Assuming a verb's lexical meaning has the same effect on English and Korean speakers, we expect the effect of IC verbs to be observed within each of the four groups, regardless of typological differences and discourse types. Across groups, we predicted an interaction between language and discourse type, namely, significantly less non-iconic causal continuations in Korean than in English in narrative discourse, but no language differences in conversational discourse.

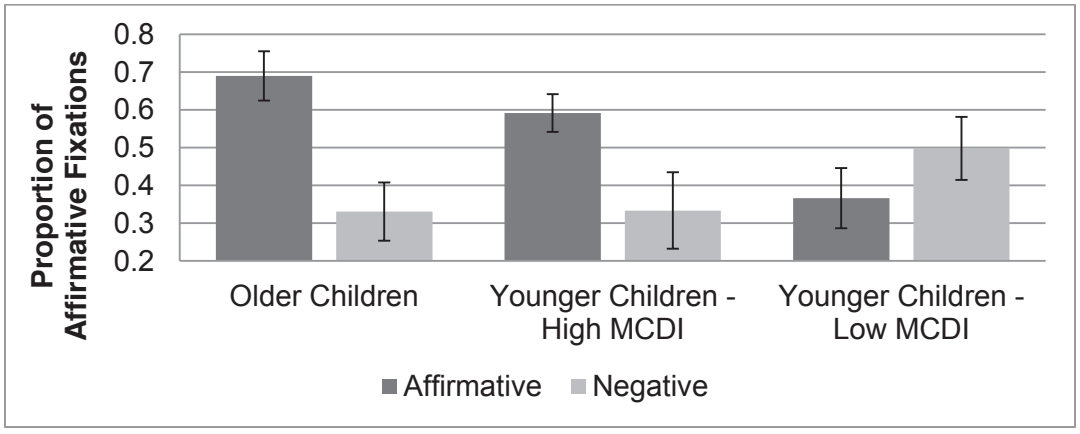
We used mixed-effects logistic regression to analyze the results of the experiment. Our hypothesis was confirmed: We found an IC effect in all four groups (main effect, $b = 1.81$, $z = 6.53$, $p < .001$) and an interaction between language and discourse type ($b = -1.51$, $z = -3.83$, $p < .001$). More specifically, (i) the effect of IC verbs in English was replicated in Korean, (ii) Korean speakers produced significantly less non-iconic causal continuations than English speakers in narratives, but no difference was found in conversations between the two languages, and (iii) Korean speakers produced non-iconic causal continuations significantly less in narratives than in conversations while English speakers behave the same across discourse types. In conclusion, our experiment provides evidence that typological differences between languages can play a crucial role in shaping speakers' discourse expectations in narrative discourse.

Young children's comprehension of negation

Tracy Brookhyser, Roman Feiman, & Jesse Snedeker (Harvard University)
tbrook@wjh.harvard.edu

Negation is a crucial test case for understanding incremental semantic interpretation, because its compositional representation is inconsistent with expectations generated by lexical items (see e.g. Panizza, 2012). Early processing studies found negative sentences were initially interpreted like their affirmative counterparts, raising the possibility that early predictive processing is associative (Kaup, Lüdtke & Zwaan, 2006; Fischler, et al., 1983). However, subsequent studies with richer discourse context found adults can rapidly integrate negation into their sentence interpretation (Nieuwland & Kuperberg, 2008; Tian, Breheny & Ferguson, 2010). No such result exists with children, and recent evidence suggests children 3 and younger do not process negation incrementally (Nordmeyer & Frank, 2013). We used the visual world paradigm to examine online sentence comprehension. Children heard a story, with corresponding pictures, which established a discourse context for both the negative and affirmative sentences (e.g. DW plans to color two stars but is interrupted), then heard the critical sentence and response prompt (e.g. "*DW didn't color/colored one of the stars. Which one was it?"*). We previously found adults show incremental interpretation of negation in these contexts. Adapting this method to children, we found older children (n = 16, *M* = 42 mos) and younger children (n = 28, *M* = 31 mos) performed differently (Fig 1). In our online measures, we analyzed the proportion of fixations to the affirmative target (e.g. the colored star) for the time between VP offset and NP onset. In older children, a mixed effects model found solely an effect of polarity ($p = .01$). In younger children, there was an interaction between vocabulary, measured by MCDI, and polarity ($p = 0.01$). There was a marginal effect of polarity among younger children with higher MCDI scores (p = .07). In offline measures, older children's picture selection was above chance in affirmative (77%) and negative (86%) trials, but younger children performed at chance in affirmative (60%) and negative (52%) trials, despite good performance in unambiguous practice trials (81%). Overall, results suggest 2-year-olds have difficulty resolving competition from two similar referents, regardless of polarity, though those with higher vocabularies may do somewhat better. 3-year-olds do not appear to have particular difficulty interpreting negation. Critically, we find no evidence that initial interpretation is associative (e.g. early affirmative fixations in both conditions). In this context, negative utterances are interpreted as quickly and accurately as affirmatives, suggesting incremental compositional processes in both cases.

Figure 1



Saying it the way you mean it: Suprasegmental cues to non-literal meaning

Georgia Simon & Kristen Syrett (Rutgers University)
geosim@ruccs.rutgers.edu

Introduction and Summary: In addition to its role in helping to disambiguate syntactic structure or mark focus [5-7], prosody can also encode aspects of non-literal meaning such as verbal irony. However, previous studies have produced mixed results, because they have conflated speaker negativity, sarcasm, and irony, relied on spontaneous speech, or situated target utterances in heavily biased contexts [1-4]. We show that speakers in a controlled production study produce suprasegmental cues to indicate whether or not what is said is what is meant, which differ reliably from speaker perspective on the asserted content, even in the absence of a previous discourse context or shared common ground. What's more, listeners appeal to these cues to determine whether the speaker intends the literal, veridical meaning or not.

Methods: Participants (n=16) were instructed to produce a series of two-line utterances like(1)-(2) in a way that would express a positive or negative attitude towards the topic.

1. Predicate of Personal Taste (PPT) n=4 x 2 versions each
   A: Carl says that's Joe's new girlfriend over there. B: She seems lovely.
2. Content Word n=4 x 2 versions each
   A: They say Diane brought a date to the party. B: Yeah, it was Dudley.

Half of the test items included a Predicate of Personal Taste (PPT), for which the positive version expressed speaker commitment to the Truth of the assertion, and the negative version expressed speaker commitment to its Falsity (i.e., NOT(she is lovely)), flouting Grice's Maxim of Quality (1). The other half included a copular construction with a content word in predicative position, where the positive version simply delivered an assertion, while the negative version encoded an additional layer of meaning: speaker disapproval/negative perspective (2). Minimal pairs were created across the two sets of test items based on number of syllabus and stress pattern (e.g, *lovely/Dudley*). Test items (n=16) were pseudorandomized with controls (Horn scale weak v. strong term, n=4; declarative v. interrogative, n=6; metalinguistic negation, n=8). Sentences were excised and the stressed syllable of the target words evaluated for key values. Two blind coders also coded each test sentence as having a falling (declarative L%) or non-falling contour. In a complementary perception study, participants listened to the excised PPT sentences and indicated whether or not what was said was what was meant.

Results

Production

Item type	% falling contour	V Duration (ms)	Intensity (dB)	Max F0 (Hz)
PPT-pos (T) (1)	100.0%	127.9	71.5	247.5
PPT-neg (F) (1')	92.2%	**146.5**	**68.1**	**153.4**
Content-pos (2)	100.0%	118.2	72.7	227.6
Content-neg (2')	96.9%	127.7	70.0	160.1

All bold acoustic values distinct from minimal pair members at p<.001 (ANOVAs)

Perception (*Is what is said what is meant?*): Average d' score: 2.75, 90.3% correct overall

Ironic, non-literal meaning exhibits a suprasegmental signature distinct from positive, veridical asserted content, and distinct from negative speaker perspective on the topic, which is in turn retrievable by hearers tasked with identifying speaker meaning.

References

[1] Bryant (2010). *Discourse Processes.* [2] Bryant & Fox Tree. (2002). *Metaphor and Symbol.* [3] B & F T. (2005). *Lang and Speech.* [4] Cheang & Pell. (2008). *Speech Comm.* [5] Price et al. (1991). *JASA*. [6] Speer et al. (1993). *JML*. [7] Syrett et al. (2015). *Journal of Linguistics*

Friday, March 14
Poster Session II Abstracts

Delaying verb production changes what matters in subject-verb agreement

Arielle Allentoff & Heidi Lorimor (Bucknell University)
hml003@bucknell.edu

Within the literature on subject-verb agreement production, one current debate is whether agreement is a "forward" or "backward" process. In forward models (e.g., Solomon & Pearlmutter, 2004; Gillespie & Pearlmutter, 2011), agreement values are generated as the noun phrase is processed, and are then transmitted to the verb. In backward models, like the Memory Retrieval Model (Badecker & Kuminiak, 2007), the relevant agreement features are retrieved as the verb is being planned. Other models of agreement, like the Marking and Morphing Model (Eberhard, Cutting, & Bock, 2005), would allow for either forward or backward agreement computation.

Within Marking and Morphing, two factors are primarily involved in agreement production: the notional number of the noun phrase and the grammatical number specifications of the morphemes within that noun phrase. If agreement is calculated in a forward direction, we would predict no difference in the relative influence of grammatical and notional information, regardless of the timing of the verb, as agreement values would be calculated at the noun phrase and retained until use. However, within a backward model, the relative contributions from grammatical and notional information should vary, as grammatical information from the morphemes would fade quickly (Sachs, 1967), while notional information from the message would remain accessible over time.

To date, much of the evidence for forward or backward models has come from investigations of word order effects (e.g., Haskell & MacDonald, 2005), which have often involved differences in linear order and/or syntactic structure. This makes it hard to determine whether the patterns found in verb agreement are due to timing or to other effects. Our goal was to investigate how timing affects the relative contributions of grammatical and notional information to subject-verb agreement by manipulating the distance to the verb, while holding constant the syntactic structure and order of the subject relative to the verb.

We selected 18 sets of experimental preambles that manipulated notional number from Humphreys & Bock (2005) and paired these preambles with adverbs. In a Sentence Completion Task, native English speakers saw an adverb (*always*) and listened to a preamble (*the gang on the motorcycles*). Their instructions were to repeat the preamble and complete the sentence using the adverb given. Participants were divided into two groups. The "verb early" group was instructed to use the adverb, but not right away, causing them to produce the verb right after the preamble. The "verb later" group was instructed to repeat the preamble, then use the adverb, and then to complete the sentence. This served to delay the verb. For example:

> Verb early: The gang on/near the motorcycles is/are *always…*"
> Verb later: The gang on/near the motorcycles *always* is/are…"

Logistic regression showed a significant interaction between verb position and notionality; participants who used the verb later were more likely to use plural verbs in the notionally plural condition than in any other condition. This finding provides evidence that agreement is calculated backwards, as the relative influence of notional number increased over time. This result also highlights the role of working memory in agreement computation and suggests that subject-verb agreement is influenced not only by the notional and grammatical properties of the subject, but also by when speakers have to make a decision about which verb to use.

Semantic relatedness and semantic integration in subject-verb agreement errors

Darrell J. Penta & Neal J. Pearlmutter (Northeastern University)
d.penta@neu.edu

Numerous sentence completion experiments have found that subject-verb agreement error rates are greater for subject noun phrases (NPs) containing a mismatch between the grammatical number of head and local nouns (Ns) relative to their corresponding matched controls, especially when singular heads appear with plural-marked local Ns. Although conceptual factors like the notional number of some kinds of Ns and NPs have been shown to influence this mismatch effect, another line of research has explored how agreement is affected by meaning properties not related to number. For example, Barker, Nicol, and Garrett (2001; BNG) reported greater mismatch effects for subject NPs in which head and local Ns were semantically related (viz., category coordinates or synonyms) versus unrelated. Separately, Solomon and Pearlmutter (2004; SP) demonstrated that increasing semantic integration increased mismatch effects, where semantic integration was operationally defined as the extent to which the constituent elements of a to-be-uttered phrase were conceptually tightly linked.

Relatedness and integration are distinct conceptual properties, in that the former is concerned with a relationship between two Ns irrespective of context, and the latter is concerned with the relationship between two Ns within the context of a particular utterance. This implies that these factors can be separately manipulated; but there is also a clear potential connection: Related Ns might be implicitly considered integrated by virtue of their being related. In addition, SP proposed that the mechanism underlying semantic integration effects on agreement might be changes in timing of lexical availability during planning, with more integrated Ns planned closer together in time, creating more interference between their number specifications; and this same mechanism could be responsible for effects of semantic relatedness on agreement, as might be predicted by a typical spreading-activation account of lexical priming created by semantic relatedness in word recognition.

To begin to investigate the relationship between integration and relatedness, the current experiment used 24 NP PP subject NPs like (1), manipulating local noun number, semantic integration [by varying the preposition linking the head and local Ns (1a&c vs. 1b&d)], and semantic relatedness [by varying the local N (1a&b vs. 1c&d)]. Norming surveys confirmed these manipulations and also revealed a link between relatedness and integration, in that the related versions were rated as more integrated than the unrelated versions. The critical stimuli were combined with 72 fillers to create 8 counterbalanced lists, and 244 participants (each saw a single list) read aloud and completed each stimulus to form a complete sentence.

Data from 196 of these participants have been analyzed with ANOVAs on (arcsine- and un-transformed) agreement error rates. Related versions yielded reliably larger mismatch effects than unrelated versions. Integrated versions produced numerically (but not significantly) larger mismatch effects than unintegrated versions, and integration did not interact with relatedness. The relatedness results correspond to BNG's and indicate that relatedness, broadly construed, is sufficient to increase error rates. We will discuss further implications of these results for agreement mechanisms, as well as the results of a second experiment that separately manipulates distinct types of semantic relationships: category coordination (e.g., *church-mosque*), used by BNG; and critical property or attribute relations (e.g., *apple-seed*), which are often involved in creating semantic integration.

(1) a. The canoe with the weathered oar(s) (Related - Integrated)
b. The canoe near the weathered oar(s) (Related - Unintegrated)
c. The canoe with the weathered flag(s) (Unrelated - Integrated)
d. The canoe near the weathered flag(s) (Unrelated - Unintegrated)

Barker, J., Nicol, J., & Garrett, M. (2001). *Journal of Psycholinguistic Research*, *30*, 91-114.
Solomon, E. S., & Pearlmutter, N. J. (2004). *Cognitive Psychology*, *49*, 1-46.

Yo pienso, *tu piensas*: Crosslinguistic attraction effects in agreement comprehension

Sol Lago (University of Maryland), Diego Shalom (Integrative Neuroscience Lab), Mariano Sigman (Integrative Neuroscience Lab), Ellen Lau (UMD) and Colin Phillips (UMD)
sollago@umd.edu

Previous studies have found that English speakers are susceptible to agreement attraction errors (e.g. *The key to the cabinets are on the table*) when comprehending subject-verb dependencies [1,2,3]. These effects have been attributed to a retrieval mechanism [3,4], in which comprehenders use morphological and syntactic cues on the verb to probe the contents of memory, sometimes resulting in the retrieval of nouns (e.g. *cabinets*) that are syntactically unlicensed but match the verb in number [3,5,6]. However, few studies in comprehension have addressed whether attraction is affected by the degree to which the morphology of a language encodes agreement information. Here we compare attraction profiles in English with Spanish, a language where agreement morphology is more marked and functionally more significant. Across three self-paced reading studies, we find that Spanish speakers are indeed more sensitive to agreement violations, but that the size of attraction effects is surprisingly similar across languages (Figure 1). However, we show that attraction profiles in Spanish but not English vary as a function of the type of verb (auxiliary vs. main verb). To capture these results, we propose that agreement engages both retrieval and predictive processes, and that retrieval is uniformly implemented across languages, but number predictions are stronger in morphologically rich languages.

The present studies tested the non-local agreement attraction effect previously found in English [3,5,7]. We manipulated the number of a relative clause (RC) head and whether the RC subject and verb agreed in number (e.g. *The note(s) that the girl writes/*write during class cheered her friend up)* in a 2 (singular/plural) × 2 (grammatical/ungrammatical verb) design. Experiment 1 (*n*=32) used main verbs within the RCs. Replicating previous English findings [3,4], we found significant reduction in the reading time disruption in the post-verbal region of the ungrammatical conditions when the attractor was plural, and no effect of attractor number in the grammatical conditions, consistent with the idea that retrieval is only engaged when an error signal in the input is encountered. Experiment 2 (*n*=32) used Spanish auxiliary verbs, where singular and plural verb forms are more closely matched in length, as in English (e.g., *va/van a escribir, 'is/are going to write'*). Here we found symmetrical attraction effects: facilitation in the ungrammatical conditions, and disruption in the grammatical ones when the attractor was plural. Finally, in Experiment 3 (*n*=32), we showed that the disruption in the grammatical conditions is not obtained with English auxiliary verbs, while still observing robust attraction in the ungrammatical conditions.

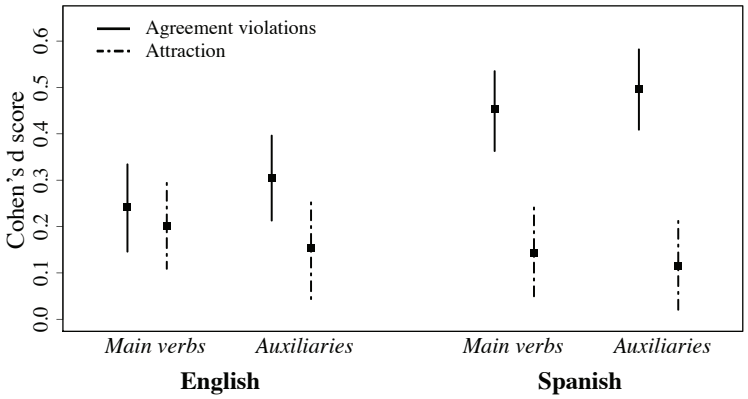


Figure 1. Effect sizes across languages using a standardized scale. Squares show mean effect sizes and bars represent 95% CIs. Data for main verbs in English was obtained from [3]

[1] Pearlmutter et al. (1999); **[2]** Nicol et al. (1997); **[3]** Wagers et al. (2009); **[4]** Lewis & Vasishth (2005); **[5]** Staub (2010); [6] Dillon et al. (2013); **[7]** Clifton et al. (1999).

Impact of phonological distance on lexical ambiguity resolution in people with and without aphasia

Maria Ivanova (National Research University Higher School of Economics (NRU HSE)), Anna Laurinavichyute (NRU HSE), Olga Dragoy (NRU HSE), Anastasia Ulicheva (Hong Kong University), & Svetlana Kuptsova (Center for Speech Pathology and Neurorehabilitation)
mivanova@hse.ru

It has been demonstrated that phonological length of utterance, which directly impacts the number of lexical items that must be activated to comprehend a sentence, is one of the crucial factors underlying comprehension difficulties in aphasia. Still it remains unclear whether individuals with different types of aphasia are all susceptible to the effect of phonological length. Further, it is unknown if the increased phonological length only negatively impacts sentence processing. We aimed to test the efficiency of online lexical ambiguity resolution depending on length of phonological materials intervening between the ambiguous word and the disambiguation region in healthy controls and people with fluent and non-fluent aphasia.

Participants (36 controls, 16 with Broca-type aphasia, 8 with Wernicke-type aphasia) were presented with spoken stories containing a lexically ambiguous word and a visual display while their eye movements were monitored. To investigate the impact of phonological length on ambiguity resolution we designed two conditions (short and long) that varied in the phonological distance between the presentation of the ambiguous word and the disambiguation point. All linguistic stimuli were in Russian. Auditory stimuli included: two introductory sentences ('A technician was getting ready for the repair works for an hour. Eventually he found a screw'), a clause with an ambiguous word ('Then he fixed the *crane*'), an extended modifier present only in the long condition ('with permanently loud and awfully annoying') and a disambiguation region ('*leaking water*'). In Russian two possible equally frequent meanings of 'crane' include - 'faucet' and 'engineering device', although the introductory sentences were designed to bias the interpretation to the target meaning, which the ambiguity was later resolved to. The accompanying visual panels included four drawings representing two meanings of an ambiguous word and two foil referents. To assess off-line comprehension at the end of the story participants were asked to answer a question about the relevant meaning of an ambiguous word.

We used a logit regression (Barr, 2008) to analyze the difference in proportion of fixations between the target and the competitor meanings of an ambiguous word in the disambiguation region. The main effect of phonological distance for all groups was found, with target advantage over competitor being smaller in the short condition compared to the long condition. Also an interaction between factors of group and time was revealed: although target advantage increased over time for all groups, participants with fluent aphasia demonstrated smaller increase compared to controls, but did not differ from participants with non-fluent aphasia, who did not differ statistically from the controls. As for off-line comprehension, the effect of phonological length was marginally significant in individuals with fluent aphasia, with the long condition being more difficult for them. In general, persons with aphasia performed significantly lower compared to controls in the off-line task, with individuals with fluent aphasia exhibiting the greatest number of errors.

The results show that the effect of phonological length is universally manifested in all tested groups and favors ambiguity resolution. We interpret lower target advantage in the short condition in the disambiguation region as a result of greater interference from competitor activation, while with additional time given in the long condition the biased target meaning became more activated than the competitor, facilitating ambiguity resolution. Also the data show that persons with aphasia are able to process lexical ambiguity on-line, albeit individuals with fluent aphasia do it less effectively. The observed mismatch between relatively spared on-line and impaired off-line processing is coherent with previous studies attributing aphasia comprehension difficulties to reduction in general processing resources rather than to aphasia-specific deficits in on-line linguistic processing (Dickey & Thompson, 2009; Thompson & Choy, 2009).

Antecedent contained deletions in native and non-native sentence processing

Oliver Boxell, Claudia Felser (University of Potsdam) & Ian Cunnings (University of Reading)
boxell@uni-potsdam.de

Antecedent-contained deletion (ACD) structures contain elliptical gaps that are contained within their own antecedent, as illustrated in (1).

(1) The scout was taught to [$_{VPi}$ catch every/the snake that John did [__]$_i$ in the outback]

The covert movement operation of Quantifier Raising (QR) is thought to help prevent interpretation ('infinite regress') problems from arising when trying to reconstruct the elided VP at the point of the gap. In sentences like (1), QR serves to raise the quantified phrase (QP) *every snake that John did/was* (but not its definite counterpart, *the snake that John did*) out of the VP containing it, thereby removing the gap from within its antecedent.

Using eyetracking-while-reading, the present study investigates whether encountering a quantified (as opposed to a non-quantified DP) object facilitates the processing of a VP gap further downstream, as was previously observed in a self-paced reading study by Hackl et al. (2012). We presented both native and German non-native speakers of English (*n*=32 each) with sentences such as (1) and (2), manipulating both the determiner (*every snake* vs. *the snake*) and the auxiliary preceding the gap so as to alter antecedent size (*was* vs. *did*) in a 2x2 design.

(2) The scout was [$_{VPi}$ taught to catch every/the snake that John was [__]$_i$ in the outback]

The QR account predicts that gap processing should be facilitated in the QP relative to the DP conditions, modulated by antecedent size if QR applies locally. From the point of view of anaphoric or “pointing” approaches to VP ellipsis (e.g. Martin & McElree 2008), neither the antecedent’s quantificational status nor its complexity should affect processing at the gap.

While both groups took longer to read quantified than non-quantified object noun phrases when these where initially encountered, only the native speakers also showed significant effects of quantification at and following the gap region, along with effects of antecedent size. Unlike what was found by Hackl et al. (2012), coming across the VP gap initially triggered longer reading times for the quantified conditions at the region following the gap (e.g. *in the outback*). The expected QP facilitation effect was restricted to the native speakers’ rereading times. Here we found a Quantification X Antecedent Size interaction [F_1=5.76,p<.05; F_2=5.61,p<.05], with the longest reading times seen in the "long VP, non-quantified" condition. We also found significantly more regressions into all regions preceding the gap for non-quantified antecedents.

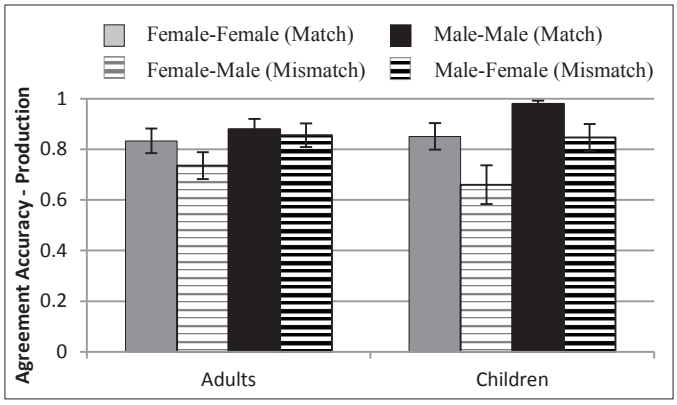
The non-native group, in contrast, only showed main effects of antecedent size at and following the gap region, with the long VP conditions (2) eliciting longer reading times, and triggering more regressive eye movements, compared to the short VP conditions (1). No main effects of, or interactions with, quantification were observed. Taken together, our results indicate that ACD resolution is affected by the quantificational status of the antecedent VP's object in native but not in non-native sentence processing. The native speakers' reading-times are consistent with a QR account and suggest that QR is triggered, or reconfirmed, when the VP gap is encountered. The non-native pattern indicates that more complex VPs are more difficult to reconstruct than simpler ones. The observed complexity effects are unexpected from the point of view of non-reconstruction approaches to VP ellipsis (e.g. Martin & McElree 2008). Taken together, our results suggest that more than one possible ACD resolution mechanism is available in principle, with native but not non-native speakers using a QR-based strategy.

References: Hackl, M. et al. (2012). Quantification and ACD. Evidence from real-time sentence processing. *Journal of Semantics,* doi:10.1093/jos/ffr009. Martin, A. E. & McElree, B. (2008). A content-addressable pointer mechanism underlies comprehension of verb-phrase ellipsis. *Journal of Memory and Language* 58, 879-906.

Gender agreement errors in learner English: Evidence from production and comprehension

Lucia Pozzan (University of Pennsylvania), Inés Antón-Méndez (University of New England), & John C. Trueswell (University of Pennsylvania)
lpozzan@sas.upenn.edu

While agreement phenomena are ubiquitous across the world's languages, individual languages differ in how these agreement relationships are implemented. For example, Romance languages typically display local gender agreement between a possessive pronoun and a local *possessum* (E1), while English possessives agree with non-local *possessors* (E2), and Chinese displays no overt gender features on possessives (E3). Here, we investigate whether language learners show a tendency to implement agreement locally by matching gender features of possessive pronouns with those of local possessums. If such a tendency exists, L1 and L2 learners of English might display higher errors rates when possessors and possessums mismatch (E4) rather than match in gender (E5), occasionally resulting in productions like E6. In comprehension, gender mismatches might result in referential uncertainty.



Results & Discussion: *Production:* 23 adult L1-Chinese learners of English and 27 English monolingual children (Mean Age: 4;0, Range: 3;0-5;4) were prompted to produce sentences containing possessive phrases (E4-E5). The gender of the possessor (female/male) and gender match/mismatch between possessor and possessum were manipulated. Significantly more errors arose in mismatch than match conditions ($p<.001$), indicating a tendency for learners to implement agreement locally by matching gender features of possessive pronouns with those of local possessums. This tendency was particularly marked in child learners (interaction: $p=.03$).

Comprehension: 19 adults and 17 children (Mean Age: 3;8, Range: 3;1-4;7) from the previous study participated. Learners' online and offline interpretation of possessive pronouns was measured as they acted out commands while their eye-movements were recorded. The manipulations were the same as in the previous experiment. There was a sharp contrast in the performance of the two groups. Adult learners performed well, unaffected by the experimental manipulation: upon hearing the possessive (e.g., "*his* little"), they looked at the target referent and/or the family competitor, with no consideration of the gender competitor, indicating a correct preference to interpret "his" as agreeing with the possessor; act-out errors were virtually nonexistent. In contrast, children produced significantly more errors in mismatch than match conditions ($p=.03$), indicating referential uncertainty when the local nouns mismatched in gender.

Taken together with previous results with Spanish and Italian adult learners[1] of English, the present findings indicate that language learners, regardless of age and native language background, have a tendency to establish agreement within a local domain—a tendency that needs to be counteracted as language-specific processes are automatized; such automatization seems to be achieved earlier in comprehension than production.

Examples: **E1.** [Maria$_{FEM}$ [$_{VP}$ ama [$_{DP}$ suo$_{\mathbf{MASC}}$ padre$_{\mathbf{MASC}}$]]]; **E2.** [$_{TP}$ Maria$_{\mathbf{FEM}}$ [$_{VP}$ loves [$_{DP}$ her$_{\mathbf{FEM}}$ father$_{MASC}$]]]; **E3.** [$_{TP}$ Maria$_{FEM}$ [$_{VP}$ ai [ta$_{\mathbf{MASC/FEM}}$ de baba$_{MASC}$]]]; **E4.** Possessor-Possessum Mismatch: *John* gave the flower to his *sister*; **E5.** Possessor-Possessum Match: *John* gave the flower to his *brother;* **E6.** *John^{i} gave the flower to *her^{i}* *sister*

References: [1] Antón-Méndez, I. (2010). Whose? L2-English speakers' possessive pronoun gender errors. *Bilingualism: Language and Cognition*, 14, 318-331.

Processing at the semantic and syntactic interface in learners of Spanish: evidence from ERPs

Carla Contemori, Patricia Romàn & Giuli Dussias (Penn State University)
cuc29@psu.edu

One recent approach to second language (L2) comprehension claims that individuals prioritize semantic cues, to the exclusion of morphosyntactic cues, when comprehending their L2 (Clahsen & Felser, 2006), and moreover, that this constitutes a fundamental difference with native (L1) comprehension, where syntactic cues should take priority. In the present study, we test if and how L2 learners of Spanish process structures that entail both semantic and syntactic information compared to native speakers. To investigate this, we have chosen constructions in which a different form of the Spanish copula (*ser/estar*) is selected in relation to the semantic properties of the noun in locative constructions (events vs. objects): in Spanish, locative predicates about objects require *estar en*, while those relating to events require *ser en*, both translated as ‘‘to be in’’ in English. We presented four sentence types in which “object/event + to be” constructions were either correct or incorrect (La silla esta / *es en la cocina - *The chair is in the kitchen;* La fiesta es / ??esta en la cocina -*The party is in the kitchen*). We measured ERPs and recorded grammaticality judgments.

We recruited 24 native speakers of Spanish and two groups of L2 Spanish speakers (24 beginners and intermediate and 18 advanced learners) to investigate how the locative construction that links semantics and syntax is processed at different levels of proficiency.

In line with previous studies on locatives in Spanish L1 (Sera, Gathje, and Pintado, 1999; Leone-Fernandez, Molinaro, Carreiras, Barber, 2011), the results of the grammaticality judgment for the native speakers showed that while the ‘‘object + ser + en’’ was considered strongly ungrammatical, the ‘‘event + estar + en’’ combination was perceived as unacceptable to a lesser degree. For these same participants, ERP recording time-locked at the onset of the critical word “en” showed a larger P600 for the *ser en* predicate when the subject is an object than when it is an event (e.g., La fiesta es en la cocina vs. *La silla es en la cocina). This P600 effect is consistent with a syntactic repair of the defining predicate when it does not fit with the adequate semantic properties of the subject. On the other hand, for the *estar en* predicate with event subjects compared to object subjects, the findings showed a central-frontal negativity between 500-700 ms, that we speculate might show the detection of an anomaly that is syntactically reparable.

Data collected with L2 Spanish speakers using grammaticality judgments show that beginners are significantly less accurate than native speakers in all conditions, while advanced speakers only differ from the natives in the event+ser and event+estar condition.

For the ERPs, advanced L2 Spanish speakers show a similar pattern as the native speakers: (1) a P600 response to the “object noun + ser” violation more central and frontally distributed, and (2) a central-frontal negativity between 500-700 ms for the “event noun + estar” violation, while beginner learners did not show any effect in the ERPs in the time-windows under analysis.

The results indicate that the beginners have not yet made a distinction between the use of Ser and Estar in locative construction, as shown by the behavioral and ERPs data. From the behavioral results, L2 learners showed a native-like pattern on the use of the object+estar verb, but not for the event+ser. However, the ERPs data showed a similar response for both sentence types. Our ERP and behavioral results suggest that L2 speakers might start by perceiving *estar* as the “default” copula in locative constructions at the lower levels of proficiency, showing more sensitivity to morphosyntax and frequency of occurrence than pure semantic cues (object vs event), contrary to (Clahsen & Felser, 2006). Furthermore, our ERPs measurements seem to suggest that in advanced learners the processing of locatives with event+ser is very similar to that of native speakers, despite their lower off-line accuracy.

Complement coercion as aspectuality: Evidence from SPR and fMRI

Yao-Ying Lai, Cheryl Lacadie, Todd Constable, Ashwini Deo, Maria Piñango (Yale University)
yao-ying.lai@yale.edu

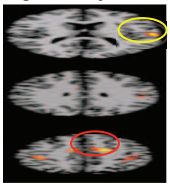
Research shows that the increased processing cost associated with complement coercion, as in (1) *John began the book*, is observable only when combining an entity-denoting complement with an aspectual verb (AspV) (e.g. *begin, finish*), but not with a psych verb as in (2) *John enjoyed the book*[1,2]. The **Dimension , mbiguity** (DA) analysis[3], a lexical-semantic analysis of AspVs captures the distinction. On this analysis, AspVs do not select for events but for structured individuals (entities that can be totally ordered on an axis along one dimension). And AspVs in turn encode *a set of functions* along a range of ontological dimensions – spatial, temporal, informational, and eventive. Thus, to interpret (1), the parser must extract a dimension from the complement that is compatible with the dimension-specific functions lexically stored in AspVs. In (1), the complement "book" offers more than one dimension: informational & eventive, leading to two mutually exclusive interpretations: the subject is construed as an initial *constitutive* part: "*The story about John was the first story in the book"* or an *agent*: *"John began doing something with the book*." Following previous generalizations about processing cost, this analysis predicts two sources of cost: **(1) exhaustive activation of the verb's lexically encoded functions**[4] and **(2) resolution of ambiguity created by immediate composition demands (i.e., *dimension extraction* from complement)**[5]. We test those predictions via self-paced reading (SPR) and investigate their neurological basis using fMRI. Previous findings associate complement coercion composition with three distinct cortical regions: Wernicke's area[6], vmPFC[7], and BA45[8]. The DA analysis reconciles these observations. It predicts AspVs to recruit Wernicke's area for **cost 1** and the left frontal cortex for **cost 2**. We compare AspVs, *Enjoy-type* psych verbs (EnjoyV), and *Love*-type psych verbs (LoveV) as controls[9]:

[*Lady Gaga*]$_1$ [***started***$_{(AspV)}$ **/ *preferred***$_{(EnjoyV)}$ **/ *loved***$_{(LoveV)}$]$_2$ [*this CD*]$_3$ [*of*]$_4$ [*American*]$_5$ [*pop hits.*]$_6$

, cceptability tests (ratings, 1-5 scale) show LoveV (*M*=4.80(0.22)) > EnjoyV (*M*=4.31(0.52)) = AspV (*M*=4.13(0.75)). With the same materials, **SPR** (N=28) shows significantly longer reading times for AspVs over EnjoyV and LoveV ($p < .05$) *at window 4* and *5*. This replicates previous observations that complement coercion composition is slow to develop[10], and that its processing effects are only observable for AspVs[1,2]. It also confirms the "control" nature of LoveV condition[9]. In the **event-related fMRI** (N=15, 50 Ss/condition), each sentence is segmented into two events, Event 1: [Subj-V] "*Lady Gaga started*" Event 2: [complement…] "*this CD…*", each isolating one hypothesized process.

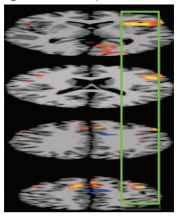
Results: At Event 1 (Fig.1), results show a preferential recruitment of **Wernicke's area (B, 40) and bilateral B, 7,6,24** for AspV over EnjoyV. At **Event 2** (Fig. 2), AspV preferentially recruits **the left lFG (B, 44,45,47) and left insula** over the control LoveV.

Fig. 1 Subj-Verb



The DA analysis captures these findings as follows: **B, 40** activation (yellow) at Event 1 reflects the verb's *exhaustive lexical activation* (arising from the dimension functions in AspV, absent in EnjoyV). BA7 (red), an area recruited for spatio-temporal tasks, is also recruited for AspVs, suggesting conceptualization of structured individual. The **LIFG & insula** activation (green) at Event 2 reflects in turn the process of *determining the dimension* along which the complement axis is structured.

Fig. 2 Complement...



Altogether, our results support the Dimension Ambiguity analysis as it captures a distinction between predicates (Aspectual vs. Psych) in a linguistically principled manner. It supports processing and brain architectures that capture the compositional independence of meaning and that distinguish lexically-related from compositionally-related semantic processes.

[1]Katsika *et al.,* 2012, [2]Utt *et al*., 2013, [3]Piñango & Deo, 2012/2013, [4]Shapiro *et al*., 1989, [5]Frazier & Rayner, 1990, [6]Piñango & Zurif, 2001, [7]Pylkkänen & McElree, 2007,[8]Husband *et al.* 2011, [9]Pustejovsky1995, [10]McElree *et al.* 2001.

Timing of lexical activation in determiner–adjective–noun phrase production

Amy DiBattista & Neal J. Pearlmutter (Northeastern University)
dibattista.a@husky.neu.edu

Determiner–adjective–noun phrase production (Det–Adj–N; e.g., *the green dress*) involves activation of a lemma and lexeme for both an N and an Adj. Schriefers (1992, 1993) investigated lemma activation order in similar Dutch NPs using picture–word interference: Semantic interference for Ns and Adjs at the same early SOAs suggested overlapping activation timing for N and Adj lemmas. However, the lemma data were obtained in separate experiments, and there was no basic phonologically related condition to generate lexeme activation data (unlike Schriefers, Meyer, & Levelt, 1990). The current Experiment 1 used picture–word interference to investigate two possible activation orders in NP production: role driven, in which the N is activated first, as the primary content element in the NP; and linear-order driven, in which the Adj is activated first, matching the ultimate order of production. Experiments 2 and 3 investigated lexical activation in the production of isolated Ns and Adjs.

The stimuli were 144 pictures from two N categories (6 shapes, e.g., triangle; and 6 clothing articles, e.g. dress), with attributes from two Adj categories (6 colors, e.g., green; and 6 patterns, e.g., striped) applied. Participants were first familiarized with the objects and (separately) the attributes to be described, along with their intended labels. During testing, participants described the pictures one at a time, with a distractor that was semantically related, phonologically related, or unrelated to the picture's N or Adj label. In relation to the target Ns and Adjs, semantic distractors were from the same category but were phonologically unrelated, phonological distractors had the same onset and nucleus phonemes but were semantically unrelated, and unrelated distractors were both semantically and phonologically unrelated. All distractors were selected from outside of the target set. Distractors were visually presented at three positive and three negative SOAs (-200, -150, -50, +100, +200, +250 ms).

In Experiment 1, participants produced Det–Adj–N descriptions. Reaction time from the picture's appearance to the response onset was analyzed with planned comparisons of the unrelated versus the semantic and phonological conditions at each SOA, separately for N and Adj targets. The N comparisons showed semantic interference at the -150 ms SOA (marginal at the +200 ms SOA). The Adj comparisons showed semantic interference at the -200, -150, and +200 ms SOAs (marginal at the +250 ms SOA). There were no phonological distractor effects. In Experiments 2 and 3, participants produced isolated nouns (e.g., *dress*) or adjectives (e.g., *green*), respectively (with correspondingly limited training). Experiment 2's Ns showed semantic interference at the -200 and -150 ms SOAs. Experiment 3's Adjs showed marginal semantic interference at the -150 and +250 ms SOAs, and phonological facilitation at the +200 ms SOA.

Experiment 2 replicated Schriefers et al.'s (1990) pattern for isolated N lemmas; and Experiment 3 suggests that for isolated Adjs, lemma activation precedes lexeme activation, as it did for Schriefers et al.'s Ns. Experiment 1 suggests overlapping N and Adj lemma activation, with Adj lemma activation initiated earlier. This points to a linear-order driven timing of activation: Lemmas were activated in their ultimate production order. An utterance's word order may therefore be known during lemma processing, which would contradict the top-down-only information flow specified in some production models (e.g., Bock & Levelt, 1994). Alternatively, the attributes may have been more salient than the objects, leading to earlier activation of the attributes' concepts and, consequently, their lexical items. In addition, Experiments 1 and 3 showed surprising late semantic interference, attributable to task effects. The three experiments will be discussed in terms of syntactic influences (driven by role or linear position assignment) and conceptual influences (driven by message-level or stimulus properties) on lexical activation.

References

Bock, K., & Levelt, W. (1994). In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 945–984). San Diego: Academic Press.

Schriefers, H. (1992). *Cognition*, *45*, 33–54.

Schriefers, H. (1993). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 841–850.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). *Journal of Memory and Language*, *29*, 86–102.

Effects of animacy and noun-phrase relatedness on processing of complex sentences

Matthew W. Lowder & Peter C. Gordon (University of North Carolina at Chapel Hill)
matt.lowder@unc.edu

Previous work suggests that syntactically complex object-extracted relative clauses (ORCs) are easier to process when the head noun phrase (NP1) is inanimate and the embedded noun phrase (NP2) is animate (e.g., *The article that the senator skimmed…*) compared to the reverse animacy configuration (e.g., *The senator that the article accused…*)[1-5]; further, some work suggests that processing differences begin as early as NP2 (before the reader has encountered any verbs)[1]. This pattern has been explained using an expectation-based framework, under which readers use the animacy of NP1 to form predictions about how the sentence is most likely to unfold, with processing difficulty at NP2 reflecting violation of those predictions[1]. However, the nature of this comparison makes it difficult to pinpoint the source of the effect, given that the ORCs that are compared differ with respect to the animacy of NP1, the animacy of NP2, and the embedded verb. A previous eye-tracking experiment that manipulated the animacy of NP2 while holding the rest of the sentence constant demonstrated that readers experienced local processing difficulty stemming from the need to integrate an inanimate subject with an action verb (e.g., *article accused*); however, this greater difficulty did not contribute to the more global difficulty associated with the complex syntax of an ORC[6]. Taking a similar approach, the current pair of eye-tracking experiments investigated effects of NP1 animacy on complex-sentence processing while holding the rest of the sentence constant.

Experiment 1 compared sentences like those in (1), which were adapted from previous materials[1,5]. Our results showed the typical ORC-SRC processing difference at the RC region and the matrix verb (1a vs. 1b). Comparison of the two ORCs showed greater difficulty for ORCs with an animate (1b) compared to an inanimate (1c) NP1. This difference emerged in regression-path duration on NP2 (i.e., before any verbs had been fixated), $F_1 = 6.93$, $p <. 02$; $F_2 = 4.32$, $p < .05$, as well as in total time on NP2, $F_1 = 10.79$, $p <. 005$; $F_2 = 8.34$, $p < .01$.

Experiment 2 compared ORCs like those in (2), where the only word that differed across conditions was NP1. NP1 varied in animacy and either was semantically related (2a & 2b), or unrelated (2c & 2d) to NP2. Analysis of regression-path duration at NP2 revealed a significant interaction, $F_1 = 6.00$, $p < .02$; $F_2 = 5.08$, $p < .05$, such that the Animate-Arbitrary condition (2c) was more difficult than the Inanimate-Arbitrary condition (2d), but animacy had no effect when NP1 and NP2 were related (2a & 2b). In addition, analysis of regression-path duration at the matrix verb revealed a significant main effect of animacy, $F_1 = 13.66$, $p < .005$; $F_2 = 10.05$, $p < .005$, such that the Animate condition was more difficult than the Inanimate condition, reflecting differences in the semantic reversibility of NPs between these conditions.

As readers begin processing an ORC, they attempt to establish meaningful relations between NP1 and NP2 before encountering any verbs. When NP1 and NP2 share a tight semantic link, readers use this information rather than animacy to form an initial representation of the two referents participating in the same event, and then easily proceed to the verbs of the sentence. If there is no straightforward semantic link between NP1 and NP2, then readers use information about the animacy of the two NPs to help form an understanding of how the two referents may be interacting in the event. This leads to easier processing when NP1 is inanimate and NP2 is animate compared to when both NPs are animate.

Experiment 1	Experiment 2
1a. The senator that criticized the journalist was..	2a. The mayor that the senator criticized was…
1b. The senator that the journalist criticized was..	2b. The bill that the senator criticized was…
1c. The article that the journalist criticized was…	2c. The waitress that the senator criticized was…
	2d. The recipe that the senator criticized was…

[1] Gennari & MacDonald (2008), *JML*. [2] Mak et al. (2002), *JML*. [3] Mak et al. (2005), *JML*. [4] Traxler et al. (2002), *JML*. [5] Traxler et al. (2005), *JML*. [6] Lowder & Gordon (2012), *JML*.

Effects of animacy and semantic relatedness during sentence processing: An ERP study

Mariana Vega-Mendoza, Martin J. Pickering & Mante S. Nieuwland (University of Edinburgh)
m.vega-mendoza@sms.ed.ac.uk

How do different forms of semantic knowledge impact our understanding of an unfolding sentence? Some researchers have argued that semantic memory organization (in particular, semantic relatedness) overrules early effects of sentence plausibility only if an incoming word fits coarse-grained semantic expectations such as animacy (Paczynski & Kuperberg, 2012; henceforth P&K). To address this issue, we investigated effects of semantic relatedness on processing of sentences containing plausibility violations or animacy violations, using materials based on P&K. Nineteen English native speakers read passive sentences (RSVP with 300/200 on/off, post-sentence yes/no plausibility judgment) with the sentential subject NP (see example sentence in Fig. 1a) belonging to a plausible-animate control condition (1) or to one of four implausible conditions: (2) related-animate, (3) unrelated-animate, (4) related-inanimate or (5) unrelated-inanimate. Relatedness was indexed with Latent-Semantic Analysis and items were pre-tested on plausibility to match P&K (31 items per condition, 90 fillers to balance plausibility).

ERP analyses focused on 350-450 ms (N400) and 700-900 ms (P600) time windows, using 2x2 repeated measures ANOVAs with factors animacy and semantic relatedness tested at sixteen electrode sites (divided in 4 topographical regions with Anterior/Posterior and Hemisphere distributions). Only statistically significant findings are reported here.

Inanimate NPs elicited larger N400s than animate NPs. Moreover, semantically unrelated NPs evoked larger N400s than related ones, but no interaction was found (Fig. 1b). With regards to the P600 window, we found a two-way interaction involving relatedness and AP distribution: related NPs elicited larger P600s at anterior regions than unrelated NPs.

Our results partially replicate P&K's: Whereas the largest N400 effects were also found for animacy violations, we obtained an additive effect of semantic relatedness, rather than the interactive pattern reported by P&K with only a relatedness effect for animate NPs. The smaller N400s for related NPs irrespective of animacy suggests that semantic relatedness facilitated processing incoming NPs regardless of the fulfillment (or lack thereof) of animacy expectations.

We repeated the study with Spanish-English bilinguals (N=17), whose first language has grammatical marking for animacy, to see whether L2 processing in this group relies more strongly on animacy cues than on semantic relatedness due to possible modulation from the L1. Results from this group indicate that bilinguals indeed show larger N400s for inanimate NPs but not for unrelated NPs.

Sample sentence:

The prescription for the mental disorder was written by the…

	Mean accuracy for yes/no plausibility judgments (%)
(1) *psychiatrist* ———	91
(2) *schizophrenic* -·-·-·-	81
(3) *guard* – – – –	90
(4) *pill* ·····	94
(5) *fence* – – –	99

Fig. 1a *Sample sentence with critical words per condition.*

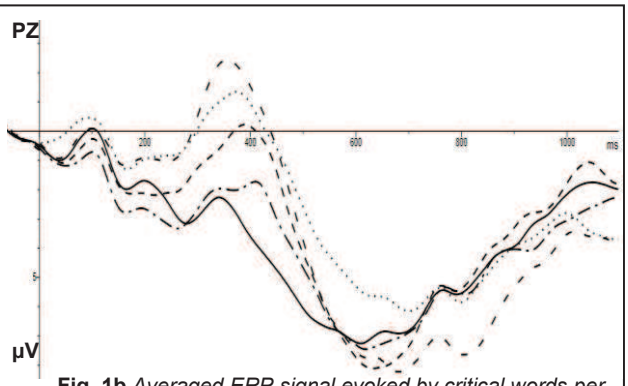


Fig. 1b *Averaged ERP signal evoked by critical words per condition at PZ.*

Paczynski, M., & Kuperberg, G. R. (2012). Multiple influences of semantic memory on sentence processing: Distinct effects of semantic relatedness on violations of real-world event/state knowledge and animacy selection restrictions. *Journal of memory and language*, *67*(4), 426-448.

A common neural basis for syntactic and non-syntactic conflict-control

Nina Hsu (University of Maryland – College Park), Susanne Jaeggi (University of California – Irvine), & Jared Novick (University of Maryland – College Park)
ninahsu@umd.edu

To understand how people process and understand sentences, it is important to consider how non-syntactic cognitive functions contribute to parsing and interpretation. One way to investigate this is to study common neurobiological mechanisms that are shared by syntactic and non-syntactic cognitive procedures. We focus on cognitive control—the ability to adjust thoughts and actions when confronted with information-conflict during processing. When parsing sentences, cognitive control may engage to capture the correct interpretation of language input after initial misanalysis [1]. For example, patients with damage to lateral prefrontal cortex (PFC) fail to revise early parsing decisions following syntactic ambiguity [2], a deficit that is coupled with cognitive control impairments in non-syntactic domains: they also show exaggerated conflict effects in the Stroop task and in item-recognition tasks when recently seen but irrelevant memoranda must be ignored [3]. Likewise, neuroimaging results show co-localized recruitment of PFC regions (e.g., posterior left inferior frontal gyrus, or pLIFG) when healthy people interpret spoken ambiguities and complete Stroop-conflict trials [4]. Thus, neural resources supporting cognitive control functions in memory may also support the resolution of syntactic conflict, but two unresolved issues remain. First, what counts as "co-localized" activity across tasks? Group analyses afford statistical power, but single-subject analyses provide increased anatomical sensitivity for better functional specificity [5]. It also remains untested whether multi-voxel pattern analyses (MVPA) could complement traditional univariate findings *within* a study. We tested the commonality of conflict-control mechanisms across syntactic and non-syntactic domains, hypothesizing common recruitment of LIFG in conditions with relatively high cognitive control demands. Twenty subjects completed four tasks while undergoing fMRI. In Stroop, they indicated the ink color of color-terms while ignoring word meaning (e.g., BLUE in red ink). In the Recent-Probes task, participants indicated whether a letter-probe was among a memory set. Occasionally, the item was from the prior but not current set, creating conflict relative to non-recent trials also eliciting a "no" response. In the 3-back memory task, subjects viewed single words sequentially and indicated if the current one matched the word three trials ago. Recently presented "lures" in nearby positions (e.g., 4-back) created conflict between familiar yet irrelevant memoranda. Finally, participants read garden-path sentences ("While the thief hid the jewelry sparkled brightly") to assess syntactic conflict-control versus a comma-disambiguated version ("While the thief hid, …"). We compared each task's conflict condition to its baseline, and then conducted single-subject analyses by identifying active voxels in Stroop within pLIFG. In these individually-defined regions, the other tasks showed reliable or marginally reliable conflict effects (3-back: t=2.57, p=0.02; Recent Probes: t=1.99, p=0.06; Syntactic Ambiguity: t=3.15, p=0.006). We did not find co-localization in other regions, and similar results emerged both when identifying voxels using a different task and when using an anatomical LIFG region of interest. These univariate results suggest that *within subjects*, common LIFG regions may selectively mediate cognitive control in language and memory; they also complement work advocating single-subject approaches [6]. Further, we used MVPA to examine whether conflict detection elicits a common brain state that can be reliably classified within tasks. Preliminary data suggest above-chance within-task classification accuracy in LIFG (Stroop: 43%, 3-back: 53%; chance=33%). We are currently testing if a classifier can reliably predict conflict states across distinct tasks. This result would suggest common conflict-induced mind/brain states consistent with a domain-general theory of cognitive control that can inform neurobiological theories of sentence processing.

References: [1] Novick et al., 2005, *CABN;* [2] Novick et al., 2009, *Cogn. Neuropsych.;* [3] Hamilton & Martin, 2005, *CABN;* [4] January et al., 2009, *JOCN;* [5] Fedorenko et al., 2011, *PNAS;* [6] Fedorenko et al., 2012, *Brain & Lang*.

Effects of Experience and Expectations on Adaptation to Dialect Variation in Noise

Kodi Weatherholtz, Abby Walker, Shannon Melvin, Adam Royer and Cynthia G. Clopper
(The Ohio State University)
kweatherholtz@ling.ohio-state.edu

Several factors affect processing of dialect variation in good listening conditions, including recent experience with a dialect (Dahan et al., 2008; Floccia et al., 2006), long-term experience with local varieties (Sumner & Samuel, 2009), and global expectations due to dialect priming (Hay et al., 2006). This study examines effects of recent and long-term dialect experience and global expectations on the processing of dialectally-varying sentences in noise.

We build on work by Clopper and Bradlow (2008), which showed that Mid-Atlantic, Northern, and Southern dialects of American English are less intelligible in noise than standard-sounding General American English (GenAm), independent of listeners' own dialect. We tested whether intelligibility of Mid-Atlantic speech *selectively* improves after familiarization (Exp1, N=35), implicit priming (Exp2, N=33), and jointly priming and familiarizing listeners to this dialect (Exp3, N=38). We also tested whether listeners showed a selective benefit for their own dialect. *Selective benefit* means exposure to one dialect improves processing of *only* that dialect.

The experiments involved an exposure-test format. The test phase was identical across experiments, but the exposure phase varied. At test, participants listened to 96 unique HP SPIN sentences mixed with speech-shaped white noise at -2dB SNR and were asked to transcribe each sentence. Sentences were produced by 6 speakers (3 M, 3 F) from each of four dialect regions: GenAm, North, South, and Mid-Atlantic. Test sentences were divided into two blocks, with each block immediately preceded by an exposure phase. In Experiment 1, the exposure phase comprised listening to short stories produced by either Mid-Atlantic or GenAm talkers, thus familiarizing participants to one or the other dialect immediately before test. In Experiment 2, the exposure phase comprised brief visual presentation of place names associated with the Mid-Atlantic or GenAm regions (e.g., New York vs. Columbus). The exposure phase in Experiment 3 combined the familiarization stories (Exp1) and associated place names (Exp2).

Mixed Poisson regression models predicted the number of content words correctly identified in each sentence by the dialect of the test talker, exposure block (Mid-Atlantic vs. GenAm), exposure block order, trial number, and participants' native dialect (North vs. Midland/GenAm). Replicating Clopper and Bradlow's (2008) results, we found significantly lower intelligibility for Mid-Atlantic, North, and Southern speech relative to GenAm, independent of participants' native dialect (i.e., no effect of long-term experience). However, Mid-Atlantic intelligibility was *selectively* improved following familiarization to and associative priming of this dialect (Exps 1 and 2), with these benefits emerging late (in an interaction with trial number; see Figure 1). There was no exposure effect in Experiment 3, possibly due to a local ceiling effect.

These sentence processing results indicate that intelligibility of dialect variation in noise is selectively affected by recent experience with and global expectations about this dialect. This selective benefit is modest, though, relative to the overall intelligibility deficit for non-standard dialects, suggesting weaker mappings for non-standard variation, regardless of experience.

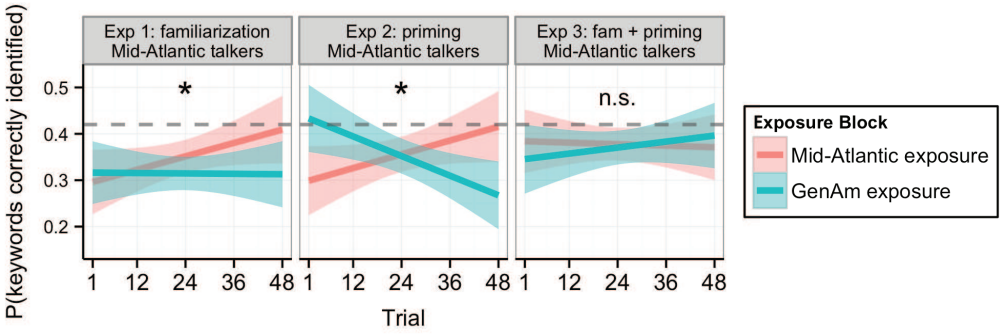


Figure 1: Accuracy for Mid-Atlantic speech by trial, exposure block, and experiment.

132 **Locality in filler-gap dependencies: Evidence from extraposition**

Brian Dillon (UMass Amherst)
brian@linguist.umass.edu

The distribution of locality (and anti-locality) effects in long-distance syntactic dependencies has been influential for theories about the memory architecture of the parser [1]. Here we examine the role of locality in filler-gap comprehension by looking at English relative clause (RC) extraposition (ExRC). ExRC displaces an RC filler to the right of its base position (i.e. a *gap-filler* configuration). We investigated ambiguous ExRC constructions as in (2). In (2) ExRC places the RC filler in a VP-adjoined position that is structurally closer to the higher NP *ex-wife,* but which maintains the relative temporal/linear recency of the potential gap sites at *ex-wife* and *soldier*. This allows us to tease apart the effects of temporal recency and structural proximity on gap-filler dependencies. Using offline ratings and online reading times, we show that the recency preference for ambiguous RC attachment is present only when the RC is adjacent to its nominal host. In contrast, ambiguous ExRC gap-filler dependencies show no preference for the more recent gap or the more structurally prominent gap. These results are incompatible with theories of filler-gap locality effects stated over linear [2] or structural distance alone [3]. Instead, we argue that these results support theories that attribute locality effects in filler-gap dependencies to interference and decay in working memory [1]. We present an implemented computational model that supports this claim. The English results contrast with previous findings on German ExRC [4], which shows a strong high attachment preference in three-site ambiguities for ExRC. We suggest that this difference may reflect cross-linguistic differences in the role of recency in RC attachment in the two languages [c.f. 5].

In a 2x2 design we crossed attachment height (LOW versus HIGH attachment) with the position of the RC (IN-SITU vs. EXTRAPOSED). We used the gender of a reflexive pronoun to disambiguate attachment height:

(1) IN-SITU: ***Recently, I met the ex-wife of the American soldier [who poisoned herself/himself with dangerous chemical agents.]***

(2) EXTRAPOSED: ***I met the ex-wife of the American soldier ___ recently [who poisoned herself/himself with dangerous chemical agents.]***

Experiment 1 (n = 20) used offline naturalness ratings with a 1-7 Likert scale (7 = most natural) to determine the difficulty associated with high and low attachment for both RC positions. The results revealed a significant interaction of attachment height and RC position ($F1(1,19) = 6.6$, $p < 0.05$; $F2(1,23) = 3.8$, $p < 0.07$). Planned pairwise comparisons reveal that this effect was due to significantly higher ratings for low attachment over high attachment for in-situ RCs (4.2 vs 3.8; $t(19) = -2.1$, $p < 0.05$). The pattern was reversed for extraposed RCs, which showed a nonsignificant trend towards a preference for high attachment (3.6 vs. 3.8; $t(19) = 1.4$, $p < 0.2$).

Experiment 2 (n = 40) used self-paced reading to investigate the online processing of these ambiguities. Participants were recruited using Amazon's Mechanical Turk, and the experiment was administered using the online experimental platform IbexFarm. Analysis of residual reading times revealed a significant interaction of extraposition and locality in the post-disambiguation region ("*with*") ($F1(1,39) = 8.3$, $p < 0.01$; $F2(1,19) = 7.5$, $p < 0.05$). Planned comparisons revealed this interaction was driven by a significant effect of attachment height for in-situ RCs (HIGH: 59ms vs LOW:-13ms; $t(39) = 3.7$, $p < 0.001$), with no such effect for extraposed RCs (HIGH: 16ms vs LOW: 14ms; $t(39) = 0.2$, $p = 0.9$). No other significant effects were found.

A model of these results using an ACT-R parser shows that these effects reflect **i**) an activation boost for the high NP due to PP postmodification and **ii**) brief, rapidly-decaying activation of the local NP. These two factors combine to produce a short-lived locality effect, which does not survive over an intervening adverb in the EXTRAPOSED conditions.

[1] Bartek, Lewis, Vasishth & Smith. (2011). *JEP:LMC* [2] Gibson. (1998). *Cognition*. [3] O'Grady. (1997). *Syntactic Development*. [4] Hemforth, Konieczny, & Scheepers. (2000). In *German Sentence Processing*, Hemforth & Konieczny (eds). [5] Hemforth, Konieczny & Scheepers (1994). *Tagundsband KNOVENS.*

Semantics overrides syntax in the processing of gap-filler dependencies in Chinese

Shukhan Ng (University of Texas at San Antonio), and Nicole Y. Y. Wicha (University of Texas at San Antonio, University of Texas Health Science Center at San Antonio)
Contact email: shukhan.ng@utsa.edu

In a gap-filler dependency, the gap precedes its filler in linear word order. Gaps are structural positions that have little semantic content until linked to fillers. Thus, gap-filler completion is motivated primarily by fulfilling interpretation requirements, i.e., supply semantic content to a sentence element. This semantic requirement can compete with syntactic requirements like structural economy for ambiguity resolution. The present study employs gap-filler processing in Chinese to investigate fulfilling competing semantic and syntactic requirements in sentence comprehension. The relevant Chinese construction is shown in (1). It contains a subject gap and can be analyzed as a main clause, subject clause (both are right-branching analyses), or pre-nominal relative clause (RC). In a right-branching analysis, the gap may co-refer with a noun in the segment that follows *friend* or is interpreted contextually. In a RC, *friend* is the filler. The right-branching analysis is overwhelmingly preferred when the clause-initial verb *wait* is received (Ng & Fodor, 2011), likely due to structural economy. Here we determine if Chinese speakers adopt the RC analysis at *friend*. This revision would violate several structural principles (e.g., Minimal Attachment and Minimal Revisions), and would show a preference for filling the gap early despite the greater structural complexity of the RC.

Results of the oral gated sentence completion study (N = 20) showed that the right-branching analysis was more preferred for the 3 shorter word strings (*wait*, *wait guest*, *wait guest de*) before *friend* in (1) but the RC analysis was more preferred for the string in (1) as a whole. Moreover, when strings in (1) were disambiguated as a subject clause (SC) by the main clause that follows *friend/reply*, as in (2), self-paced word-by-word reading results (N = 36) showed significantly longer reading times for several words in the main clause, the main clause as a whole, and the last word of the sentence in (2c), compared to the other unambiguous sentences. (The overt subject in (2a,b) and the inanimate noun *reply* in (2d) block the RC analysis.) These findings suggest that the RC structure is built at the noun *friend* in strings like (1) and disambiguation in favor of the subject clause parse produces a reading slowdown. More importantly, only when a suitable filler (*friend* vs. *reply*) is available will the more complex RC analysis be adopted. Overall, sentence processing in Chinese may be initially driven by structural economy but semantics can override syntax and triggers a reanalysis.

(1) *e* Dengdai keren de pengyou… (*de*: a marker for noun modifiers)
wait guest DE friend
a. Main Clause: '(Someone) waited for the guest's friend.'
b. Subject Clause: 'Waiting for the guest's friend…'
c. Relative Clause: 'The friend who waited for the guest…' ([$_{RC}$ e_i wait guest] de friend$_i$)

(2) a/b. [$_{SC}$ **Women** dengdai keren de **pengyou**/**dafu**] yijing henjiu-le.
We wait guest DE **friend/reply** already long-time
c/d. [$_{SC}$ Dengdai keren de **pengyou/dafu**] yijing henjiu-le.
wait guest DE **friend/reply** already long-time
'It has been a long time (for us) to wait for the guest's friend/reply.'

References
Ng, S., & Fodor, J. D. (2011). Use your headedness: An exercise in psycholinguistic exploitation. In H. Yamashita, Y. Hirose & J. L. Packard (Eds.), *Processing and Producing Head-final Structures* (pp. 299-321): Springer.

Structural and Non-Structural Locality Effects in Bangla Filler-Gap Dependencies

Dustin A. Chacón & Colin Phillips (University of Maryland), Mashrur Imtiaz & Sikder Monoare Murshed (Dhaka University), Shirsho Dasgupta & Mina Dan (Calcutta University)
dchacon@umd.edu

Locality biases are pervasive in language processing and syntactic theories, but it is disputed whether one type can be unified with the other (cf. Sprouse & Hornstein 2013). In particular, locality effects that impact the acceptability of filler-gap dependencies (FGD), i.e., "island effects", can only be unified with on-line locality biases if they employ the same notion of locality. Cross-language research, primarily comparisons of English with head-final languages like Japanese, challenges this approach by showing that the tail of a FGD preferentially resolves with the first verb in a sentence, even if that verb is in an embedded clause, hence structurally distant from the filler (Aoshima *et al*, 2004; Omaki *et al* 2014). This suggests that resource-induced locality effects in parsing are determined *linearly*, not *structurally* like island effects. The current study challenges this conclusion, based on 3 experiments on Bangla (studies conducted in Bangladesh), a language with both English- and Japanese-type orders. Although interpretive measures replicate the bias for linear locality irrespective of structural locality, measures of disruption suggest that active FGD completion may be more structural in nature.

In Bangla, the main verb may precede or follow an embedded clause, and *wh*-fronting creates FGDs, (1). This permits a controlled, within-language comparison of word order effects on FGD resolution, since it is possible to contrast "congruent" word orders (i.e., main verb = first verb) and "incongruent'' word orders (main verb ≠ first verb).

1. a. *Emily* ***kothae*** *ækjɔnke* *boleche* [ʂ' *je* *se* *nijeke* *bætha.koreche*] ? (congruent)
   Emily **where** someone told that she self hurt ?
   b. *Emily* ***kothae*** [ʂ' *se* *nijeke* *bætha.koreche* *bole*] *ækjɔnke* *boleche* ? (incongruent)
   Emily where she self hurt that someone told ?
   '**Where** did Emily tell someone that she hurt herself?'

Exp1 (n=24) was a Question-After-Story task (de Villiers *et al*, 1990), in which Bangla-speaking participants watched a series of videos. In each video, a character participates in some salient event, and reports on it in another location (order counterbalanced). Afterwards, an ambiguous *wh*-question like (1a) or (1b) is presented. The participant's response indicates the parse assigned to the sentence. The results show that Bangla speakers resolve the dependency with the first verb, regardless of its structural depth ($X^2 = 35.34$, $p < 0.001$). This reversal in interpretation biases as a function of word order reinforces previous findings from English and Japanese (Omaki *et al*, 2014).

Exp2 (n=36) was a Self-Paced Reading study, using a *filled-gap* paradigm (Crain & Fodor 1985, Stowe 1986). We manipulated word order (congruent 2a vs incongruent 2b), and the type of wh-phrase, such that it either conflicted with an overt object NP that preceded the first verb (argument-wh: filled-gap) or did not conflict (adjunct-wh: control). RTs showed a filled-gap effect in the congruent conditions (p=.015) but not in the incongruent conditions (p=.55). Exp3 (n=32) was a 7-point acceptability judgment task on the same sentences. Ratings showed an acceptability penalty in the 'filled gap' conditions in the congruent order (p=.019), and not in the incongruent order (p=.65), paralleling the findings in Exp2.

Thus, whereas the interpretive measure (Exp1) shows a uniform bias to interpret a wh-phrase with the first verb, irrespective of its structural depth, measures of difficulty (Exps2&3) show that blocking this possibility has differing consequences: it is costly only when the first verb is also the highest verb, This suggests either that reanalysis is cheaper when the linearly local gap is not structurally local, or that reanalysis is not even needed, because dependency formation is less 'active' when the first potential gap is linearly but not structurally local. Under either of these accounts of the contrast, we conclude that locality in on-line FGD construction is not entirely reducible to a linear bias, and hence this lowers a potential barrier to unifying accounts of locality in language processing and linguistic theories.

Online processing of parasitic gaps: Evidence from eye-tracking

Cassandra Chapman (McMaster University)
chapmc3@mcmaster.ca

This paper presents results from an eye-tracking experiment indicating that participants are sensitive to potential parasitic gap (PG) locations; namely, that they look longer at the verbal region where a potential PG could be licensed. This effect was found in both subject and adjunct PG constructions, constructions that crucially differ based on the order of the PG and its licensor or *real* gap.

Background PG constructions have been described as a repair strategy for island violations (Ross, 1967; Chomsky, 1982; Engdahl, 1983): a second syntactic gap is inserted in the structure, which is able to license the gap found in the island. For example, a single gap following *visiting* in (1a) results in an adjunct island violation. However, inserting a second gap in the structure in (1b) is grammatical; the gap in the adjunct is *parasitic* on the first or *real* gap. Notice that the PG can be replaced by a co-indexed pronoun, *him* in (1b), showing that the first gap is the *real* gap.

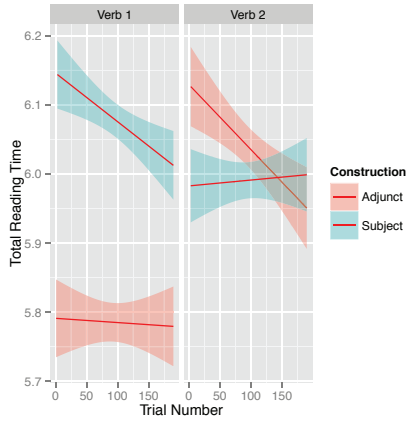
(1) a. **Island**: *Which man$_i$ did you see him$_i$ [before visiting t_i]?
b. **PG/Real gap**: Which man$_i$ did you see t_i [before visiting t_i/him$_i$]?

PGs are also available as a repair strategy for subject islands, compare examples (2a) and (2b). Crucially, (2b) demonstrates that the second gap is the *real* gap in this condition.

(2) a. **Island**: *What$_i$ did [the study to test t_i] impressively demonstrate the results$_i$?
b. **PG/Real gap**: What$_i$ did [the study to test t_i/ the results$_i$] impressively demonstrate t_i?

Therefore, subject and adjunct PG constructions differ in the structural position of the PG: the PG *follows* the real gap in the adjunct condition but *precedes* the real gap in the subject condition.

Predictions The goal of our study was to test if participants are sensitive to PGs and their licensing restrictions. We predicted longer reading times on verbs preceding gap positions. If participants are also sensitive to PG licensing restrictions, longer reading times should be observed on verbs preceding the PG than on verbs preceding the real gap. Specifically, we predicted longer reading times on Verb 2 in the adjunct condition but on Verb 1 in the subject condition.



Results *Verb 1* A linear mixed effects regression model was fitted to total reading times (the summed duration of all fixations) on the verb. Results indicate that reading times on Verb 1 were significantly longer in the subject construction in comparison to the adjunct construction ($p < 0.05$). An interaction of construction type by trial number further revealed that participants spent less time reading Verb 1 in the subject condition as the experiment progressed ($p < 0.05$; Figure 1, left panel, reading times in log units). *Verb 2* Total reading times on Verb 2 in the adjunct construction were numerically longer than in the subject construction; this effect was not significant ($p = 0.2$). Yet a construction type by trial number interaction suggested that participants spent less time reading Verb 2 in the adjunct condition as the experiment progressed ($p < 0.05$; Figure 1, right panel).

Discussion Online results indicate that reading times were longer at the verbal region where a potential PG could be licensed: Verb 1 in the subject and Verb 2 in the adjunct construction. Significant interactions of total reading time and trial number were also observed, indicating that participants began to read the verbs at a faster rate as the experiment progressed, but only if a potential PG could be licensed in this position. Therefore, the results support the hypothesis that participants are not only sensitive to PG positions but also to their syntactic licensing restrictions.

Understanding ambiguous idioms in context:
Clearing the air on compositional and noncompositional views

Nyssa Z. Bulkes and Christopher M. Grindrod (University of Illinois – Urbana-Champaign)
bulkes2@illinois.edu

It is uncontroversial to suggest that context is important in language processing. To date, little research has specifically addressed how contextual information guides comprehension of figurative language, especially with expressions having plausible figurative and literal interpretations, such as ambiguous idioms (e.g., *a slap in the face*; Papagno & Cacciari, 2010). In the psycholinguistic literature, several theories have been proposed to explain how idioms are processed. The *noncompositional view* suggests that idioms are represented and understood like long words (Bobrow & Bell, 1973; Gibbs, 1980; Swinney & Cutler, 1979). In addition, the more predictable or syntactically frozen an idiomatic phrase is, the faster the figurative meaning can be retrieved relative to the literal interpretation. In contrast, the *compositional view* suggests that like literal language, comprehenders decompose idioms into their component parts and use this semantic information to arrive at the intended interpretation (Gibbs & Nayak, 1989; Nunberg, 1978). Finally, the *hybrid view* incorporates aspects of the two previous views, suggesting the more familiar or predictable an idiom is, the easier the idiomatic interpretation can be retrieved from the lexicon (Libben & Titone, 2008; Titone & Connine, 1999).

In light of the above-mentioned theories, the present study tested younger adults' ability to comprehend idioms with plausible idiomatic and literal interpretations based on the surrounding context. All idioms were judged as being familiar and having a literally plausible interpretation based on published norms (Cronk et al., 1993; Libben & Titone, 2008; Schweigert & Cronk, 1992; Titone & Connine, 1994). In a semantic priming paradigm, participants listened to idiom-biased sentences ending in ambiguous idioms (e.g., *The couple had been together for several years so they decided to tie the knot*) or literal-biased sentences (e.g., *The friends had noticed the sail coming loose so they decided to tie the knot*). Participants were asked to make a decision whether a visually presented target related or unrelated to the idiomatic or literal interpretation was a word of English or not while reaction time (RT) and accuracy were recorded. According to the noncompositional view, priming for idiom-related targets following idiom-biased contexts should be observed relative to when literal-related targets follow literal-biased contexts. If idioms require a distinct processing mode from literal language, then priming should only be observed for idiom-related targets following idiom-biased contexts, as deviating from the figurative interpretation should require additional processing. According to the compositional view, however, priming should be exhibited for idiom-related targets in idiom-biased contexts and literal-related targets in literal-biased contexts. Further, if context plays a role, priming should only be obtained for targets when they are contextually appropriate.

Analyses of RT data using repeated-measures ANOVA and planned pairwise comparisons revealed priming for idiom-related targets following sentences where the idiom was used figuratively. Priming was also observed for literal-related targets following sentences where the idiom was used in both its literal and figurative sense. The lack of priming for idiom-related targets following literal-biased contexts suggests that idiomatic and literal interpretations are not pursued in parallel. Rather, our results support the compositional view of idiom comprehension, in that the available context guided listeners to decompose the idiom's component parts, while also directing them to activate only the appropriate figurative or literal interpretation. In the case of ambiguous idioms where the meaning can be derived from the component words (e.g., *tie the knot*), a strong context can provide a listener with enough information to make the figurative sense of an idiom as easily retrievable as the literal meaning. Moreover, it appears that context has the ability to override the dominant figurative meaning of familiar idioms in cases where this meaning is incongruous with a context biased toward the literal interpretation of the phrase.

Abstract agreement: Children's sensitivity to subject-verb agreement in comprehension does not require knowledge of specific lexical co-occurrences.

Cynthia Lukyanenko & Cynthia Fisher (University of Illinois, Urbana-Champaign)
lukyane1@illinois.edu

In sentence comprehension both adults and children rapidly integrate linguistic information at multiple levels to build meaning [1,2]. Among other cues, listeners use number-marked verbs to facilitate processing of a following subject noun phrase: they look earlier to a target picture when they hear a sentence with an informative agreeing verb than one without (*Where are the good cookies?* vs. *Look at the good cookies!;* display: 1 apple, 2 cookies; [3,4]).

What knowledge underpins this ability? Both adult and child listeners use an agreeing verb to facilitate processing even when the actual number of objects is held constant (*Where are the pretty glasses?;* display: 1 pair of glasses, 1 phone; [4,5]). This suggests that, even for 3-year-olds, agreement plays its role primarily as a part of a system governing the legal combinations of words in the language, not by conveying number meaning directly.

Agreement is typically described in terms of abstract grammatical number features. However, agreement also creates lexical-distributional patterns in the linguistic input to which children and adults are sensitive [6]. Sensitivity to either grammatical or lexical patterns can explain the existing data. On a grammatical account, number-marked verbs facilitate the processing of a subject NP with the right grammatical number features. In contrast, on a lexical-distributional account, number-marked verbs (*is* or *are*) facilitate the processing of particular word-forms with which they often co-occur. To what degree can 3-year-olds use agreeing verbs to cue a grammatical property, as opposed to a likely co-occurring word?

To investigate this question, we introduced child (n=95, mean age 38 mos.) and adult (n=32) participants to four novel count nouns. First, pictures of single novel objects (e.g., one citrus reamer) were labeled by phrases that identified the novel words as count nouns but provided no information about agreement (e.g., *Look, a keppin!*; compare: *a flour, *a glasses). Next, in a test phase, participants saw pairs of pictures that differed in the number of novel or familiar objects shown, accompanied by sentences with or without an agreeing verb (see figure). This experimental condition was compared to a control condition in which the distractor picture matched the target in number, rendering the agreeing verb uninformative in context.

Visual fixations during the sentences reveal that the processing advantage conferred by an agreeing verb extends to novel nouns. In novel noun trials, participants used the number-marked noun to find the target. But critically, the agreeing verb conferred an additional advantage in informative trials in the experimental but not the control group: participants looked longer at the target in informative than in uninformative trials in a 2 s window beginning at determiner onset.

This pattern of results suggests that, by age 3, listeners' use of agreement in online comprehension is not solely reliant on distributional learning about the co-occurrence of particular lexical items. Rather, agreement appears to be represented in terms of abstract grammatical properties.

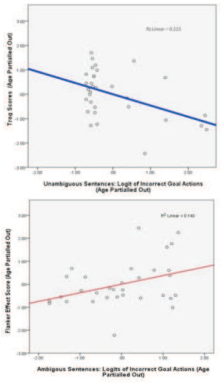
References
[1] Allopenna, Magnuson & Tanenhaus (1998) *Journal of Memory and Language* [2] Lew-Williams & Fernald (2007) *Psychological Science* [3] Lukyanenko & Fisher (2010) *BUCLD Presentation* [4] Lukyanenko & Fisher (2013) *CUNY Poster* [5] Lukyanenko & Fisher (2012) *CUNY Poster* [6] Arnon & Clark (2011) *Language Learning & Development.*

Mapping the kindergarten-path: cognitive predictors of child sentence processing

Woodard, K., Pozzan, L., & Trueswell, J. C. (University of Pennsylvania)
woodardk@sas.upenn.edu

An important issue in psycholinguistics concerns the extent to which processing capacities change across development and are influenced by the maturation of other cognitive skills. For example, children are known to differ from adults in their ability to recover from initial mis-interpretations[1] and use referential context to drive their interpretations.[1,2] 5-year-olds' eye-movements and act-out patterns show that upon hearing a temporarily ambiguous sentence (e.g., Put the cat [$_{PP1}$ on box] [$_{PP2}$ onto the pan]), they initially interpret PP_1 as the goal of movement, and fail to revise this interpretation after hearing the disambiguating PP_2. Moreover, their interpretations are not affected by referential context: differently from adults', they are not more likely to interpret PP_1 as a DP-modifier when the scene includes two referents of the same kind. Children's difficulties revising initial interpretative commitments have been linked to their immature executive functioning,[3] while lack of sensitivity to referential context might also be linked to reduced language experience.[4] To test these hypotheses, we examined the extent to which individual differences in children's sentence processing capacities correlate with differences in executive functions (conflict-monitoring/inhibition, cognitive flexibility, and working memory) and receptive language skills.

Methods: 32 monolingual 5-year-olds (Mean Age: 4;10–Range: 4;0-5;9) were administered seven tasks over two sessions. Children's conflict-monitoring/inhibition abilities were measured by their accuracy on a task in which familiar objects had to be referred to with the opposite name (e.g., sun/moon, up/down), no-go trials in a flanker task in which children were to inhibit a prepotent motor response, and differences in RTs between congruent and incongruent trials in the flanker task (i.e., flanker effect). Children's cognitive flexibility was measured by their performance on a game in which cards were to be paired together according to multiple dimensions (e.g., color and shape). Verbal working memory was assessed by a task in which children were asked to remember the last object in an array, and the standardized TROG-2 test was used as a measure of receptive language abilities. Finally, children's processing skills were measured online and offline as they acted out temporarily ambiguous and unambiguous "*put*-sentences" while their eye-movements were recorded. Ambiguity and referential context were manipulated within subjects.



Results and Discussion: As expected, temporarily ambiguous sentences were associated with significantly higher proportions of looks to the incorrect goal and higher proportions of incorrect-goal actions than unambiguous sentences, indicating that children tended to interpret PP_1 as the goal of movement and subsequently struggled to revise this initial interpretation. There was no effect of referential context and no interaction. After age was partialled out from all variables, a stepwise multiple regression analysis indicated that children's interpretative difficulties in unambiguous, but not ambiguous sentences, were predicted by their receptive grammar scores (*Figure 1*). In contrast, offline performance in ambiguous—but not unambiguous—sentences was predicted by children's flanker-effect scores (*Figure 2*). Difficulties revising initial interpretations in ambiguous sentences correlated with difficulties disregarding task-irrelevant information. These results are consistent with the hypothesis that processing abilities are linked to individual differences in general cognitive abilities. While performance in unambiguous sentences is predicted by individual differences in verbal abilities, successful interpretation of temporarily ambiguous sentences is linked to conflict-resolution skills, above and beyond the influence of age and verbal skills.

References: [1]Trueswell et al. (1999). *Cognition*. [2]Snedeker & Trueswell, (2004). *Cognitive Psychology*. [3]Novick, Trueswell, & Thompson-Schill (2005). *Behavioral Neuroscience*.[4]Kidd & Bavin (2005). *Journal of Child Language*.

Sensitivity to online encoding and retrieval interference in younger and older adults

Dave Kush, Clint Johns & Julie Van Dyke (Haskins Laboratories)
kush@haskins.yale.edu

We investigated (i) the effects of proactive interference from task-irrelevant distractors on encoding and retrieval operations during sentence processing and (ii) whether interference varies as a function of age. In a self-paced reading experiment participants were asked to maintain a list of distractor words for recall after sentence reading. Test sentences required retrieval in order to establish a dependency between a filler and a later gap.

Experiment 1 tested encoding interference by examining the impact of phonological similarity between distractor items and the filler. Previous work has observed negative effects of phonological similarity on filler-gap dependencies (e.g., Acheson & MacDonald, 2011). Here, participants were required to memorize 3 distractor words that rhymed with the filler (PhonoLoad condition), 3 unrelated words that did not rhyme (UnrelatedLoad condition), or no words at all (NoLoad condition; see 1). All memory words were controlled for lexical characteristics such as frequency, length, number of syllables, etc.

(1) **LoadType:** {won-none-run} | {term-knife-bread} | X-X-X
Test Sentence: It was the **gun** [that the actor who posed for a photo] **fired** …

We observe that both younger (n = 48, age range 18-30) and older groups (n = 42, age range 31-67) exhibit effects of encoding interference in the PhonoLoad condition at or immediately following the filler (*gun*, both *p*s < .01). Older participants (n = 42, age range 31-67) showed additional difficulty at the verb in the PhonoLoad condition only (LoadType x Age interaction; p < .01), suggesting an inability to suppress information that was salient during encoding. This is consistent with a large body of literature demonstrating that susceptibility to memory interference increases as a function of age (e.g., Healey, Hasher, & Campbell, 2013). No effects of age or condition were observed on comprehension questions, suggesting that even sensitivity to encoding interference at the gap site does not affect comprehension in either group.

We conclude that encoding interference, at least with respect to the phonological dimension, does not effect filler-gap integration, except for older adults, who are generally more susceptible to interference of all types. Older individuals are also highly susceptible to retrieval interference, as evidenced by the presence of both online and offline effects with a small sample size. In contrast, younger individuals suppress encoding interference in order to establish filler-gap dependencies, however initial data suggests that this is not possible for retrieval interference.

Acheson, D. J., & MacDonald, M. C. (2011).. *JML*, *65*(2), 193-207.
Healey, M. K., Hasher, L., & Campbell, K. L. (2013). *Psychology and aging*, *28*(3), 721.
Van Dyke, J. A., & McElree, B. (2006). *JML 55*(2), 157-166.

Skewed lexical distributions facilitate recursion learning in an artificial grammar task

Pyeong Whan Cho, Garrett Smith, & Whitney Tabor (University of Connecticut)
pyeong.cho@uconn.edu

Research on language acquisition suggests that a skewed distribution, in which a few lexical items occur more frequently in a syntactic construction than other lexical items of the same type, can facilitate generalization of the construction to new cases [1]. This learning advantage is consistent with evidence for Zipfian distribution of lexical items within syntactic constructions [2]. But Zipf's Law applies across entire language corpora [3], suggesting that the learning principle might be at work across the full grammar induction process. We therefore used an artificial grammar paradigm to see if skewed lexical distributions facilitated mirror recursion learning from scratch. We found that participants who were trained on sentences with up to three concentric phrases (Levels 1-3) with a skewed lexical distribution generalized better to Level 4 cases than participants for whom the lexical types were uniformly distributed in training. Our results provide the first evidence that skewed training helps in a recursion learning task.

Four boxes were arranged in a diamond on a computer screen with a fifth box in the middle. When participants clicked on the screen, one of the boxes changed color. The task was to click the box that would change color next. The sequence of color changes obeyed a mirror-recursive system: Root $\rightarrow$ S p, S $\rightarrow$ a S b, S $\rightarrow$ a b, S $\rightarrow$ x S y, S $\rightarrow$ x y, where a, b, x, y, and p are labels (not visible) for the four peripheral and the central boxes. 257 sentences generated by this grammar were strung end to end for a total of 1331 trials. During trials 1 through 567 ("training phase"), only levels 1-3 occurred. During trials 568 to 1331 ("test phase"), all 4 levels occurred. We manipulated two factors in the training phase: the proportion of a's and x's was either skewed (0.8:0.2) or balanced (0.5:0.5), and the order of presentation was either random or ordered such that participants saw lower levels before higher levels. 96 undergrads participated. We interpreted correct performance on predictable trials of Level 4 sentences (test phase) as evidence that participants approximated a recursive system. A mixed effects model revealed that skew enhanced recursive generalization ($p<.032$). "Starting small" was not helpful. When poor learners (ACC $\leq$ .33 on Levels 1-3) were excluded, the pattern was clearer. The same pattern held even when only the less frequent words were considered, suggesting that the benefit of skew was not wholly driven by better performance on the more frequent words.

Figure 1 shows the proportion of correct sentences (correct across all deterministic transitions) against its rank. Monte Carlo simulations suggest that the high performance of some participants cannot be achieved using a finite-state grammar of the training phase sentences (plus random noise) that responds to a Level 4 sentence based on the longest n-gram from training that matches part of that Level 4 sentence, suggesting the participants' recursive generalization.

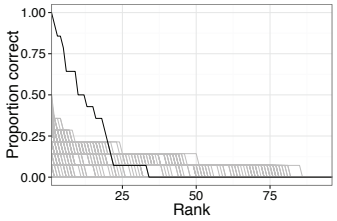


Figure 1. Proportion of correct sentences on first-instances of level 4 sentences by participant (black) and model (grey) versus proportion rank.

In sum, we find evidence for the possibility of mirror recursion learning in the lab and that skew facilitates induction of the recursive system. In contrast to [4], we did not find evidence that "starting small" enhanced generalization, possibly because the ordering schedule was not ideal. The reported recursive generalization induced from input very different from natural language is consistent with the view that syntax learning in natural language is accomplished by a general cognitive mechanism [5].

[1] Boyd & Goldberg (2009). Modern Lang. Jrnl., 93(3), 418-429. [2] Goldberg, Casenhiser, & Sethuraman (2004). Cog. Ling., 15(3), 289-316. [3] Zipf. (1935). The psycho-biology of language. [4] Lai & Poletiek (2013). Jrnl. Cog. Psych., 25(4), 423-435. [5] Perfors, Tenenbaum, & Regier (2011). Cognition, 118(3), 306-338.

The effect of syntactic category on advance planning in sentence production

Shota Momma, Robert Slevc, & Colin Phillips (University of Maryland)
shotam@umd.edu

Recent models of sentence production posit that the scope of lexical planning before utterance onset is largely a function of available cognitive resources and/or time pressure, and thus assume no architectural reason for advance planning (Ferreira & Swets, 2002; Wagner et al., 2010). However, most studies in this area are limited to the planning of nouns in simple formulaic sentences (e.g., *x is next to y*). Here we examine the effect of syntactic categories on the planning mechanism by examining active and passive utterances that change the relative ordering of verbs and arguments. We show that verbs, which impose syntactic constraints on their internal arguments, are planned before speech onset of internal argument nouns (though not before that of external arguments; Schriefers et al., 1998; but cf. Schnur et al., 2006), even under increased speech-relevant cognitive load and with increased distance from utterance onset. In contrast, we find no evidence for advance planning of nouns in non-initial position, even under less cognitive load and less distance from the utterance onset. This contrasting pattern of advance planning between nouns and verbs suggests that syntactic categories influence lexical planning above and beyond the influence of cognitive resources.

In Exp. 1 participants (n = 24) described partially colored line-drawn pictures (Fig. 1, Szekely et al., 2004), while ignoring auditorily and visually presented distractor words (SOA = -150ms). Either relevant (e.g., *chickens*) or irrelevant entities (e.g., *an apron*) were colored in the pictures. In one block participants named the object entities with or without a color adjective (e.g., "*(brown) chickens*"). In another block participants described the action in the format: *V-ing (the) (color) N* (e.g., "*feeding (brown) chickens*"), again with or without color terms. In both blocks, the distractor words were sometimes (25%) semantically related to the target nouns (e.g., target: *chicken*; distractor: *turkey*). The order of blocks was counter-balanced. Semantic similarity between target and distractor nouns delayed utterance onset (i.e., semantic interference) only when the object noun was the first word in an utterance (Fig. 2). This suggests that internal argument nouns are not planned before verb or color adjective production, even when they are linearly close to the sentence-initial verb/adj. (# of intervening words = 0-1).

Figure 1

In Exp. 2 a new set of participants (n = 18) saw the same set of pictures with verb distractors. When the patient/theme entities were not colored, participants described the action using passive sentences (e.g., *the chickens are being fed*); when patient/theme was colored, participants described the color of the object(s) (e.g., *the chickens are brown*). Robust semantic interference from verb associates (e.g., *serve*) was obtained in passive utterances, but not in color descriptions (Fig. 3). This suggests that verbs are planned before internal arguments, despite the fact that (a) participants had to make a decision between two different sentence structures across trials (unlike in Exp. 1) and (b) that verbs were relatively further away from the sentence-initial nouns (# of intervening words = 2). The null effect in the color description conditions shows that the effect is not simply due to an accidental relation between the distractors and the nouns or the pictures.

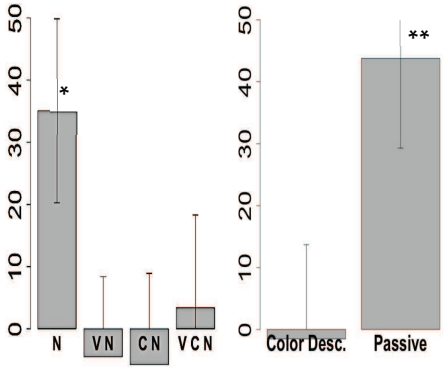


Figure 2: Noun Interference (ms)

Figure 3: Verb Interference (ms)

These results suggest that a verb's lemma is planned before the production of its internal argument nouns, but that a noun's lemma is not planned before a verb, even under the condition where a resource-based account should make the opposite prediction, based on the effect of task-relevant cognitive load and linear distance.

Verb-initial structures in Arabic: Qualitative ERP differences between sg. & pl. subjects.

R. Muralikrishnan (New York University Abu Dhabi) & Ali Idrissi (UAE University Al Ain)
r.muralikrishnan@nyu.edu

Arabic verbs generally agree with their subject in person, number and gender. For plural subjects, the verb agrees fully in the subject-verb (SV) order, whereas in the verb-subject (VS) order, verb agreement should be partial (only gender and person) if the subject is overt, full if the subject is dropped. In other words, for a VS structure with an overt subject to be grammatical, the sentence-initial verb must show singular agreement regardless of whether the overt subject is singular or plural. Note that a sentence-initial plural verb per se does not constitute a violation, provided the subject is dropped. In the present ERP study, we investigated whether the processing system is sensitive to this idiosyncratic behaviour of plural subjects in verb-initial intransitive structures with an overt subject. Our hypotheses were as follows. First, there should not be any differences in the ERPs for singular versus plural verbs sentence-initially. Second, at the position of the subject, plural subjects must show qualitatively different ERPs as opposed to singular subjects that follow a singular verb. It remains to be seen what this difference will be. Third, if the processing system initially adopts a non-anomalous reading for plural subjects, effects related to the violation must be observed at the position of the following material that conclusively signifies the anomaly.

Stimuli: intransitive sentences of the form Verb-Subject-Adverb–PP. Subject nouns: human (masculine & feminine) common nouns, singular or plural; Verb: singular or plural marked. Four conditions: 2 subject types x 2 condition types. Adverb was identical ('*yesterday*') in all conditions. Thirty-six sentences were presented per participant in each condition; all conditions were equiprobable. Stimuli were pseudo-randomised including fillers that constituted an unrelated design, and presented in a rapid serial visual presentation. Participants (34 right-handed Arabic native-speakers) performed an acceptability judgement and a probe task.

Results & Discussion: There were no effects at the position of the verb. At the position of the subject, singular subjects elicited a negativity as opposed to plural subjects regardless of condition type, and all conditions except the sg.V-sg.S condition elicited a late positivity. Whilst the two conditions with a sentence-initial plural verb would have become anomalous when the subject was encountered, it appears that the processing system did not conclude as such until the adverb (which indicates the end of the clause), at which point both anomalous conditions elicited a negativity as opposed to the acceptable conditions. Although singular subjects elicited a negativity regardless of the condition-type, those following a plural-marked verb additionally engendered a late-positivity, suggesting that the negativity for singular subjects following singular-marked verbs (and therefore grammatical) must be interpreted differently from the biphasic negativity–late-positivity pattern elicited by singular subjects following plural-marked verbs (violations). By contrast, plural subjects did not elicit a negativity, but only a late-positivity, regardless of condition-type, which can be plausibly interpreted as reflecting the syntactic integration difficulty (1) and the predictions that the processing system must make about oncoming material (2). At the position of the adverb (identical in all conditions), both violation conditions elicited a negativity, reflecting the fact that the anomaly has been finally perceived. Taken together, these results suggest that the processing system is sensitive to the idiosyncrasy of plural subjects in Arabic, and prefers to analyse them at first as syntactically difficult but nevertheless not conclusively anomalous in intransitive verb-initial structures.

References [1] Kaan et al. (2000). Language & Cogn. Processes, 15, 159–201.
[2] Frisch et al. (2002). Cognition, 85, B83–B92.

Intrusion effects on NPI licensing in Turkish: Does the parser ignore the grammar?

Aydogan Yanilmaz & John E. Drury (Stony Brook University)
aydogan.yanilmaz@stonybrook.edu

Studies of the processing of negative polarity item (NPI) licensing[2-4] have documented INTRUSION EFFECTS, where NPIs seem to be attracted by preceding but structurally illegible licensors (e.g., [*A man* [*who has no beard*]] *has* **ever left*). In ERP studies, unlicensed NPIs typically elicit P600 effects, though other negative-going responses including N400s and late left/anterior negativities have also been reported.[1] In intrusion contexts, ERP violation effects (and sentence-final behavioral rejection rates) are attenuated, raising questions about the extent to which parsing mechanisms may entertain dependency relations which are not countenanced by the grammar. One view[4] suggests that the memory/retrieval mechanisms at work in establishing licensor/NPI dependencies are responsible for intrusion (e.g., partial matching of retrieval cues). Alternatively, intrusion may result from erroneous pragmatic inferences arising from an interaction between restrictive relative clauses and negative quantifiers, which together can sometimes conspire to generate negative implicatures capable of licensing NPIs.[3]

We conducted an ERP reading study (N=22) in Turkish, where, unlike previously tested languages (English/German), NPIs typically precede licensors (a prospective dependency). We tested sentences with complement (not relative) clauses and the NPI ***kimse*** ("anybody") in matrix subject position (e.g., [NPI […embedded-Verb] main-Verb]), manipulating the presence/absence of negation as in (1A-C). Without the negation in (1A) matrix subject-NPIs are unacceptable (1B/1C). Embedded negation in (1B), however, constitutes an intervening but structurally inaccessible licensor. Though this is *not* predicted to yield intrusion effects on the erroneous negative implicature view (since, unlike like restrictive relatives, complement clauses are not expected to give rise to the contrastive implicatures which are claimed to interact with negative quantifiers), intrusion here would be compatible with the memory/retrieval view.

✓1A ***Kimse*** **Ali'nin çalıştığını söyle*medi* bana**
NPI [E.Subj. E.Verb] M.Verb-*NEG*
*1B ***Kimse*** **Ali'nin çalış*madığını* söyledi bana**
NPI [E.Subj. E.Verb-*NEG*] M.Verb
*1C ***Kimse*** **Ali'nin çalıştığını söyledi bana**
NPI [E.Subj. E.Verb] M.Verb

Unlicensed NPIs in Turkish (1C), relative to (1A), yielded a biphasic N400/P600 pattern (similar to findings in German[2]) at the main verb (after the 700 ms mark in Fig. 1, blue trace relative to black). Further, behavioral violation effects (sentence final acceptability judgments) were attenuated by the presence of the intrusive licensor. ERP reflexes of intrusion took the form of a P600-like positivity following the *embedded verb* (1B) (red in Fig.1). However, though this effect resembles part of the downstream main verb violation pattern for 1C (blue), it turns out that (1B) did not differ from additional control cases *where there was licit local licensing of an embedded clause subject NPI by the embedded negation* (not shown here). Thus, the apparent embedded verb P600 effect for (1B) can be understood to reflect (grammatically illicit) licensing at this point. Since these findings cannot be explained under the negative implicature view[4], we discuss the pro/cons[5] of explanations in terms of a memory/retrieval account.[3]

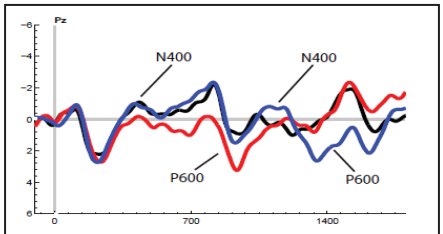


Figure 1 ERPs time-locked to the embedded verbs in (1A-C). Main verb onset at 700 ms.

References. **[1]** Steinhauer et al. (2010). Syntax, logic, and concepts in the temporal dynamics of language processing: Evidence from event-related potentials. *Neuropsychologia* 1525-1542. **[2]** Drenhaus et al. (2005). Processing negative polarity items: when negation comes through the backdoor. In: Kepser & Reis (eds.) *Linguistic Evidence- Empirical, Theoretical, and Computational Perspectives*, 145-165. **[3]** Vasishth et al. (2008). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, 32, 685–712. **[4]** Xiang et al. 2009. Illusory licensing across dependency types: ERP evidence. *Brain and Language* 108, 40-55. **[5]** Phillips et al. (2011). Grammatical illusions and selective fallibility in real-time language comprehension. In Runner, (ed), *Experiments at the Interfaces. Syntax and Semantics*, 37, 153-186.

Priming competes with syntactic anticipation, both within and across languages: Evidence from the visual world paradigm

Francesca Filiaci, Thomas Bak, Frank Keller, and Antonella Sorace (University of Edinburgh)
francesc@ling.ed.ac.uk

Humans process language with remarkable speed and accuracy. Incoming speech is assigned an interpretation in real time, updated incrementally as more input becomes available. A number of processes contribute to this efficiency. One of them is syntactic priming: a structure is comprehended more quickly when it is repeated. Secondly, anticipation processes allow listeners to predict upcoming linguistic structure based on the current input (Kamide et al., 2003).

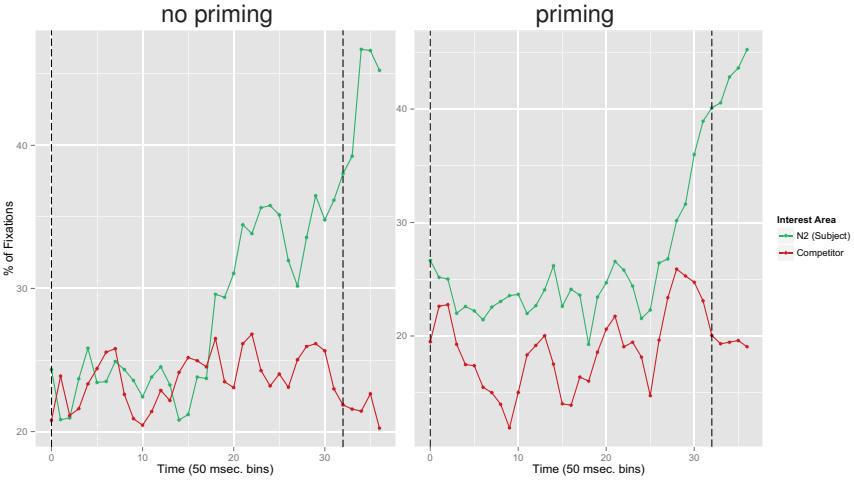
Here, we investigate how priming and anticipation interact, as the two processes have so far been studied separately. Intuitively, if a syntactic structure A is primed (occurs in the previous sentence), then it should be easier to predict A in the current sentence, as A already activated. We test this hypothesis in a visual world experiment involving both within-language and across-language priming. The experiment tested anticipation using Italian sentences such as (1):

(1) La lepre **la mangia sicuramente** la volpe.
The hare CL eats certainly the fox 'The fox eats the hare certainly'

In (1), the preverbal clitic *la* indicates an OSV structure, i.e., it makes clear that *la lepre* is an object and allows the listener to anticipate a subject.

Participants (balanced late bilinguals with L1 Italian and L2 English) were eyetracked while listening to sentences such as (1) and viewing corresponding visual scenes (here, depicting a hare, a fox, a cabbage, and a distractor). Each target (always Italian OVS) was preceded by either an OVS sentence or an SVO sentence (priming/no priming condition). The prime was either in Italian or English (same/different language condition); English OVS was realized as a passive.

Fixation proportions to the target object (here, the fox) were analyzed across the critical region (boldface in (1)). We graph the results for the different language condition, the same language condition matches closely. We find a significant anticipation effect in the no priming condition, starting half way through the critical region. In the priming condition, in contrast, the anticipation sets in significantly later but target fixations are increased at the start of the critical region.



Contrary to expectation, these results show that priming and anticipation are antagonistic: in the presence of priming, anticipation is significantly delayed, presumably because the target structure is already activated, and prediction can only take place once this activation has decayed. This holds both for within-language and across-language priming, and has important theoretical repercussions: it indicates that priming and anticipation are competing for the same memory resources, as predicted by a spreading activation model of priming (Reitter et al., 2011).

References

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research*, *32*, 37–55.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, *35*, 587–637.

Priming Enriched Meanings

Lewis Bott (Cardiff University) & Emmanuel Chemla (Institut Jean Nicod & LSCP)
Bottla@Cardiff.ac.uk

Interpreting a sentence requires considering not only the words that have been uttered, but also those that have not, that is, the *alternatives.* For instance, if a speaker says, “John’s paper was ok”, the listener can use the alternatives (e.g., “John’s paper was excellent”) to derive the meaning that while John’s paper was ok, it was not excellent. In this paper we consider how the processor derives the alternatives and enriches the basic meaning. Our approach was to apply a structural priming technique (e.g., Raffray & Pickering, 2010) to test whether enrichment can be primed. The central hypothesis was whether there is a shared procedure that is core to a range of enrichments, or whether separate procedures apply for each type.

Participants saw pairs of pictures and clicked on the picture that best described the sentences (modelled on Raffray & Pickering, 2010, and Huang, Spelke & Snedeker, 2013). Experimental trials could be *prime* trials, in which the pictures forced a particular sentence interpretation (with or without enrichment) or *probe* trials, in which either interpretation was possible. Priming occurred if interpretations of the prime sentences influenced interpretations to the probe sentences. We used three types of enrichment, each of which are arguably derived using alternatives: (1) classic scalar implicatures, *some* (*some* => *some but not all*), (2) number expressions (at least *n* => exactly *n*), (3) *ad hoc* expressions (*There is a square* => *There is a square and nothing else*). For example, *There is a square in the picture* has a weak meaning (without enrichment: There is a square and possibly other objects) and a strong meaning (with enrichment: There is square and nothing else). Figure 1 shows the example and the experimental set up. We tested within-expression priming (e.g., strong *some* prime -> *some* probe) and between-expression priming (e.g., strong *some* prime -> *ad hoc* probe).

In two experiments (N’s = 100) we observed large within-expression priming (p’s < .001) and smaller, but significant, between-expression priming (p’s < .05). In a third experiment (N = 100), we found that control expressions involving plurals were not primed by the other expressions, eliminating the possibility that assumptions of the speaker’s informativeness were responsible for the priming effects. Our findings suggest that these enrichments involve a core, shared mechanism (contra to recent experimental work by Teil et al., 2013) but that there are also large individual expression components to the decision about whether to enrich (similar to the lexical boost effect in structural priming, e.g., Pickering & Branigan, 1998). Possible mechanisms include a search for the alternatives to a speaker’s utterance, and a separate process for making particular alternatives salient.

Our study is the first to demonstrate that enrichment-with-alternatives can be primed. Although our results are broadly consistent with many accounts of enrichment (most notably Grice, 1989), there are no theories that provide precise explanations for why the likelihood of enriching the basic meaning of an utterance depends on whether an enrichment has just been made. Our data therefore provide a challenge to existing models. We hope that future work applying our priming technique to a broader range of inferences will be able to understand how enrichment-with-alternatives relates to other forms of enrichment.

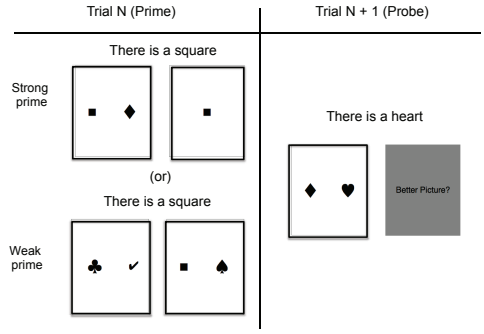


Figure 1. Participants select the picture that best matches the sentence. They see either a strong prime (with enrichment) or a weak prime (without enrichment), followed by a probe. On the probe trial, they can select either the image or the “better picture” option (see Huang et al., 2013). Priming occurs responses to the probe are influenced by the prime.

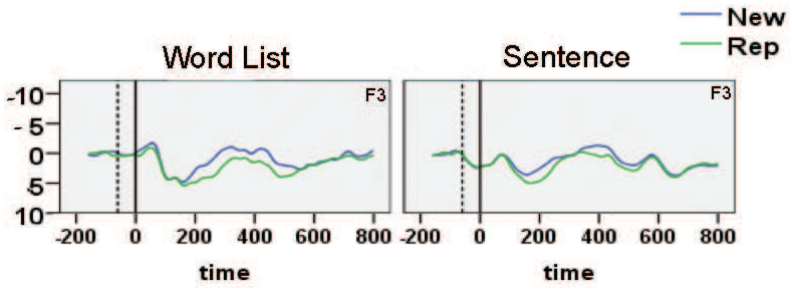
There's more to a sentence than its words: Repetition priming in sentences and lists

Giulia C. Pancani, Peter C. Gordon & Joseph B. Hopfinger (University of North Carolina at Chapel Hill) gpancani@unc.edu

Comparing how lexical factors influence the recognition of words in sentences and in lists provides a way of establishing the extent to which sentence processing affects the way that lexical cues are used. One phenomenon which is known to facilitate processing of words both in sentences and in lists is lexical repetition; this facilitation can occur for both visible and masked primes and has been observed with behavioral and neurophysiological measures.

When target words immediately follow their primes, effects of repetition are seen on event-related potential (ERP) components that are associated with both ortho-phonological (N250) and lexical semantic processing (N400) but only appear on the latter component when the target is separated from the prime [1,2,3,4]. This may be because processing of intervening words resets the state of activation in recognition units thereby eliminating the basis for early facilitation based on ortho-phonological form. Processing a visible prime may cause the target to be perceived as a separate event, again leading to the resetting of activation in prelexical processing units. Although these explanations may account for data from masked-priming studies, they are inconsistent with evidence from sentence-reading studies where eye-movement data show that early processes of word recognition are facilitated by lexical repetition even when a number of clearly visible words intervene between the prime and the target. Alternatively, repetition effects on early ERP components (N250) may vanish because reading words with the goal of understanding the sentence in which they appear causes participants to focus more on the meaning of the sentence than on the features of any single word.

The present study tests this alternative account by integrating the masked repetition priming technique with sentence reading using Rapid Serial Visual Presentation (RSVP). In Experiment 1a participants read sentences in RSVP and answered True/False comprehension questions. Embedded in each sentence were two target words which were primed by either the same word (e.g., SHOWER-shower) or a different word (e.g., CIGAR-shower); non-critical words were primed by a random string of letters. Primes were presented for 30ms, followed by a 20 ms blank screen. Targets appeared for 300ms, followed by a 200ms blank screen. Experiment 1b featured the same conditions and timing but the target words were presented in a word list and participants performed a lexical decision task on non-word fillers.



Results show main effects of repetition on the N250 [F(1,19)= 17.23, p<.001] and on the N400 [F(1,19)= 28.48, p<.001] as well as significant interactions with presentation mode (sentence vs. list) for both components (N250 : [F(1,19)= 8.07, p<.05] ; N400: [F(1,19)= 4.18, p=.05]). Follow-up analyses showed that when words were presented in a sentence the benefit of repetition was smaller, suggesting that recognition in sentence processing relies less heavily on the orthographic and lexical features of a target word than when words are presented in lists.

[1] Holcomb & Grainger (2006) JoCN [2] Duñabeitia et al., (2009) Neuroreport [3] Misra & Holcomb, (2003) Psychophysiology [4] Ledoux et al. (2007) Memory and Cognition.

Acceptability, statistical distribution, and non-usage-based mechanisms in processing null pronouns, overt pronouns, and repeated names in Brazilian Portuguese

Jefferson de Carvalho Maia (University of South Carolina), Maria Luiza Cunha Lima (Universidade Federal de Minas Gerais), Amit Almor (University of South Carolina), & Carlos Gelormini Lezama (Instituto de Neurología Cognitiva, Consejo Nacional de Investigaciones Científicas y Técnicas)
decarvaj@email.sc.edu

Referential expressions like repeated names, overt pronouns, and null pronouns have been shown to be affected by processing penalties such as the *repeated-name penalty* (RNP) [1] and the *overt-pronoun penalty* (OPP) [2]. These effects are characterized by the greater difficulty language users experience in processing repeated names and overt pronouns, respectively, in comparison to more reduced anaphoric forms when they refer to syntactically salient antecedents. Separately or jointly, such penalties have been attested in languages which significantly differ from one another with respect to their anaphoric inventories (e.g., English, Chinese). Despite their remarkably widespread scope, it is still unclear, however, whether comprehension phenomena like the OPP and the RNP reflect specific statistical regularities/acceptability in the language (e.g., [3]) or more general, non-usage-based processing mechanisms (e.g., [4]).

Brazilian Portuguese (BP) can provide important insight into this question due to an ongoing change affecting the distribution of null and overt expressions in subject position: Overt pronouns are increasingly replacing null subjects in the domain of language production [5]. Despite the shift towards lexically expressed subjects in BP, null pronouns have been recently shown in a series of on-line self-paced reading comprehension experiments [6, 7] to be more easily processed than their overt counterparts and than repeated names when their antecedents are in subject position. Still, the off-line processing of nominal and pronominal expressions in BP and, more specifically, the extent to which the documented change in statistical patterns has affected the BP speaker's metalinguistic perception of how acceptable the above-mentioned referential forms are in discourse, are unknown.

To help fill this gap, we carried out a written questionnaire study in which seventy-two native BP speakers rated the acceptability of each of the 36 items from [6]'s experiment on a 1-7 Likert scale where 1 indicated 'very bad' and 7 'very good.' The questionnaire also included 36 fillers, with pseudo-randomization for the total 72-passage set across 6 lists and with counterbalanced conditions in a 3 x 2 design with Antecedent Salience (Subject, Object) and Referential Form (Null Pronoun, Overt Pronoun, Repeated Name) as independent variables. The results were analyzed by means of repeated measures ANOVAs and revealed that, when the antecedent is in subject position, BP speakers prefer overt pronouns as anaphoric devices to null pronouns or repeated names. Interestingly, this result becomes accentuated when the reference is to the object of the previous sentence.

Thus, while null pronouns are less acceptable than overt anaphoric expressions for references to object antecedents, mirroring self-paced reading results for object conditions, they are also less acceptable for references to subject antecedents, in contrast to the reading time results available in the literature for subject conditions. This finding confirms that the RNP and the OPP occur in BP despite the shift in both use and acceptability of null subject pronouns. Anaphoric resolution seems, therefore, to be better explained as the result of a balance between processing cost and discourse function [4] rather than as a reflection of usage patterns or acceptability/sensitivity to statistical distribution [3].

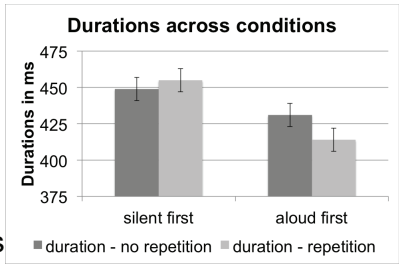
References

[1] Gordon, Grosz, & Gilliom (1993) [2] Gelormini-Lezama & Almor (2011) [3] MacDonald (2013) [4] Almor (1999) [5] Roberts & Kato (1996) [6] Gelormini-Lezama et al. CUNY2012 [7] Maia (2013).

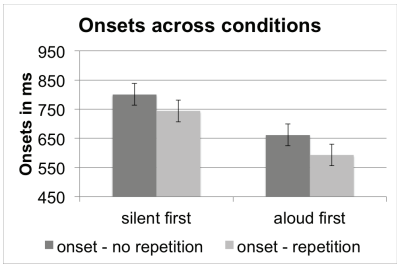
Reduction in duration depends on articulation, not simply facilitated processing

Cassandra L. Jacobs, Duane G. Watson, & Gary S. Dell (University of Illinois)
cljacob2@illinois.edu

In conversation, speakers shorten words that are given, predictable, or repeated. In this work, we investigate which aspects of the production system are implicated in prosodic reduction by exploring two hypotheses. Kahn & Arnold's (2012) Facilitation-Based Reduction Hypothesis (FRH) posits that facilitation at any level of the production system leads to shorter durations [1], [2], [3]). A second hypothesis is that reduction results specifically from articulatory representations being primed (Lam & Watson, in press). One prediction from FRH is that reduction of a repeated word will only occur if that word has actually been articulated. Kahn & Arnold (2013) tested the effects of covert articulation on reduction. Although they found a 22ms reduction in the duration of repeated words when the first mention was overtly articulated, there was only a 2ms difference when the first mention was produced as inner speech, raising the possibility that inner speech may not facilitate phonetic processing.



We sought to expand on their findings by running a more highly-powered experiment (N=55 as opposed to N=20) to investigate the role of non-articulatory facilitation in acoustic reduction. In our task, two events are presented, a shrinking event first, followed immediately by a flashing event; participants were asked to describe each event before seeing the next, e.g. "*The alligator shrinks*" and "*The frog flashes*". In the *overt* block, participants articulated both events. In the *covert* block, participants described the first event using unmouthed "inner speech", which has been argued to be phonetically, but not lexically, impoverished ([3], [4], [5]). The target word in the second event could be either new or repeated in the same trial ("*The alligator flashes*").



We measured the target *onset latency* and the target's *duration* as a function of repetition and the articulatory nature of the prime. An articulatory account like Lam & Watson's (in press) predicts reduction in duration in the *overt repeated* prime case only, but the FRH predicts reduction in both conditions. In both the overt and covert conditions, repetition led to earlier onsets ($\beta = -0.09$, SE = 0.032, $p < .01$), however there were shorter target durations *only* in the overt speech condition. The interaction was reliable ($\beta = 0.04$, SE = 0.021, $p < .05$). On the assumption that inner speech is phonetically impoverished, we conclude that the execution of a learned articulatory plan leads to reduction ([4], [5], [3]). Facilitation of onsets and preceding words, as seen also in [2], seems to be driven by general facilitation in the production system. Reduction in target durations, by contrast, requires primed articulatory representations.

References: [1] Kahn & Arnold (2012). Speaker-internal processes drive durational reduction. *JML*. **[2]** Kahn & Arnold (2013). Articulatory and lexical repetition effects on durational reduction: speaker experience vs. common ground. *LCP*. **[3]** Lam & Watson (in press). Repetition Reduction? *JEP:LMC*. **[4]** Oppenheim & Dell (2007). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*. **[5]** Oppenheim & Dell (2010). Motor movement matters: The flexible abstractness of inner speech. *Mem. & Cog*.

Eye movements during reading reflect subsequent sentence memory: Evidence from subject and object relative clauses

Brennan R. Payne, Ariel N. James, Elizabeth A.L. Stine-Morrow, & Duane G. Watson
(University of Illinois at Urbana-Champaign)
payne12@illinois.edu

Studies of sentence recall and off-line comprehension have revealed that there is considerable variability in the extent to which complete and accurate sentential representations are established in memory (Christianson et al., 2001; Gilchrist et al., 2008; Roberts & Gibson, 2002). At the same time, it is unclear whether variance in sentence memory is determined in part by initial on-line processing. Research on eye-movements during reading has revealed important mechanisms underlying the time-course of comprehension as it unfolds on-line (Rayner, 2009). In some models of sentence processing, early fixation measures are argued to be reflective of immediate interpretive processes, and as such, are more sensitive to lower-level lexically driven properties, while later-pass measures are argued to be more sensitive to higher-level "post-interpretive" factors (see Caplan & Waters, 1999; Clifton, et al., 2007). Under such models, the influence of eye-fixations on memory would be expected to occur primarily in later-pass measures that reflect re-reading and strategic encoding.

In the current study, we investigated the within-person relationship between off-line sentence recall and on-line sentence processing by adapting the subsequent-memory (SM) paradigm (whereby encoding is examined as a function of recall; Wagner & Paller, 2002) for use in sentence processing. Comparing the eye-movements of sentences that were successfully and unsuccessfully recalled would presumably reveal correlates of memory encoding.

Participants (N = 41) read subject-extracted (SR) and object-extracted (OR) relative clauses while their eye-movements were monitored. Following each sentence, participants were cued to immediately recall each sentence as close to verbatim as possible. Production was recorded and scored for recall. Mixed-effects models were fit to the data.

First-pass measures were longer for OR compared to SR sentences (an "OR cost") at the matrix verb (MV; King & Just, 1991; Traxler, 2005). Regression path duration (RPD) showed no sensitivity to syntactic complexity at the MV, but did show a large end-of-sentence slowing (i.e., wrap-up effect) that was marginally larger for OR sentences, likely reflecting a memory encoding strategy. The production data showed no overall difference in proportion of verbatim recall for SR (62%) and OR (61%) sentences. Subsequent memory effects were tested by examining whether fixation durations differed for sentences that were successfully recalled verbatim (V) compared to sentences that were not (NV). The end of sentence "wrap-up" effect in RPD was largest among sentences that were successfully recalled, but this effect was not moderated by syntactic complexity, reflecting an independent end-of-sentence integration strategy. Most notably, we found a reliable syntax x recall interaction in early-pass eye-fixation measures at the main verb, indicating that the effects of syntactic complexity differed as a function of subsequent recall. Specifically, we saw that NV sentences showed a reliable OR cost in first fixation duration and single fixation duration at the MV. Items that were successfully recalled, however, showed no OR processing cost. In sentences where the parser encounters immediate processing difficulty associated with syntactic complexity, subsequent memory for those sentences is also disrupted.

Overall, our results suggest that off-line memory effects can be reliably detected in on-line processing during sentence reading, with qualitatively distinct behavioral consequences for early and late-pass measures. Contrary to interpretive parsing models, our findings suggest that the earliest eye-movement measures can delineate subsequent memory performance, with effects occurring within the first 200-300ms.

Processing resumptives in Mandarin relative clauses: An eye-tracking study

Li-Hsin Ning, Kiel Christianson (University of Illinois at Urbana-Champaign) & Chien-Jer Charles Lin (Indiana University)
ning3@illinois.edu

Resumptive pronouns have been taken as a grammatical device that saves the ungrammaticality (e.g., Empty Category Principle in McDaniel & Cowart (1999); unextractability in McKee & McDaniel (2001)). Other than the grammaticality saving function, resumptives in general are disallowed in relativization or wh-questions. However, resumptives in English are preferred over gaps when the fillers are linearly far apart (working memory limitations and distance effect) or structurally high above (embedding effect) (Dickey, 1996, Alexopoulou & Keller, 2002, 2003, 2007). While adults accept resumptives only in unextractable sites (e.g., *the lion that Grover doesn't know what (it) ate*), children accept resumptives even in extractable positions when the resumptive pronoun is distant from the head noun (e.g., *pick up the cat that Goofy is petting (it)*) (McKee & McDaniel, 2001). In other words, resumptives can assist comprehension of longer filler-gap relations. In this study, we investigate whether a resumptive pronoun may become more acceptable and comprehensible in Mandarin when its position is farther away from the head noun and when the relative clause structure becomes more complex.

We tracked eye movements while reading Chinese sentences to investigate whether these preferences for resumptives would also occur in Mandarin relative clauses, which involve a head-final structure ("gap-filler" instead of "filler-gap" relation) with gaps more common than resumptives. We examined subject relative clauses (SRCs) and object relative clauses (ORCs) in Mandarin and lengthened the distance between gap/resumption and head noun by adding adverbials right after gap/resumption. This led to two (SRC, ORC) sets of 2 (short, long) x 2 (gap, resumption) stimuli (1-2). We also examined indirect object relative clauses (IORCs) to see if gaps and resumptives are compatible in terms of processing these more complex structures.

(1) [[**Gap**/**RP** **(ADV1)** **(ADV2)** V1 N1] DE Head]
___/?ta *teyidi* *zai xuanju* *jian* *tuidong fantanwuan* *de* *na-wei buzhang*
he purposely in campaign in push anti-corruption REL that-CL minister
V2 N2
que chuangchu *shouhuide chouwen*
but came.out bribery scandal
'The minister who ___/?he pushed the anti-corruption act has involved in a bribery scandal.'

(2) [[N1 V1 **Gap**/**RP** **(ADV)**] DE Head]
xuexiao *yaoqing* *___/ta* *duoci* *de* *na-wei* *zuojia*
school invite him many.times REL that-CL writer
V2 N2
ceng *huode* *zhimingde* *dajiang.*
ever get well-known award
'The writer whom the school invited __/him many times has got well-known awards.'

The eye-tracking results from 16 native speakers of Mandarin, analyzed with linear-mixed effects models, showed resumption did not impose a burden on processing when structure is complex (with no difference between gap IORCs and resumptive IORCs in terms of the gaze duration, total dwell time, go-past time, and regression-out rate of the interest areas (Head/V2)). Given that gaps are expected to be more acceptable than resumption, this suggests that structural complexity makes resumption more acceptable. We also found resumption facilitated processing when the gap-filler relation is lengthened in ORCs but not in SRCs, as resumptive ORCs lengthened by adding an adverbial had lower regression-out rates from V2 than their lengthened, gapped ORC counterparts, and than all the short ORCs ($p < .05$). This suggests that resumptives aid the comprehension of longer gap-filler relations in Mandarin ORCs. In sum, resumptive pronouns in Mandarin not only serve as a grammatical device to save ungrammaticality, but also become as easy to process as gaps when the structure gets complicated or when gap-filler distance increases.

Relative clause processing and competing pressures in an agreement-rich language

Manuel F. Borja (Inetnun Åmut yan Kutturan Natibu, CNMI), Sandra Chung &
mwagers@ucsc.edu Matt Wagers (UC Santa Cruz)

There is renewed interest in whether comprehension of relative clauses (RCs) reflects universal mechanisms [1-3], and how a more diverse sample of languages can bear on this issue. In most languages, a preference for subject RCs (SRs) prevails [3]. The SR preference is contentious in Chinese, which has pre-nominal RCs [4, 5, cf. 2]. A nuanced picture has also emerged in some Mayan languages, in which competing pressures to assign the 'subject' grammatical function and to license (dependent) ergative case can give rise to symmetrical extraction difficulty [3]. Here we present new evidence from the Austronesian language Chamorro, a verb-initial language of the Marianas. In Chamorro, RCs can precede or follow the head noun. In a picture-matching+touch-tracking study, Chamorro post-nominal RCs show the subject advantage but pre-nominal RCs are preferentially interpreted as object RCs (ORs). Based on touch-trajectories, inter-island variation, and error rates for unambiguous RCs, we argue this apparent OR advantage conceals the familiar pressure to link the head to a subject gap, a pressure ultimately outcompeted by the RC-internal NP's ability to agree with the verb.

Design. Chamorro is among a small group of languages with both pre-nominal and post-nominal RCs, as well as headless RCs. These three types are illustrated in the table below with the transitive verb *dengkut* ('peck'), which agrees with the subject in person/number. For transitive verbs, speakers allow both SR and OR interpretations. In contrast, passive verbs (*dinedengkut*) and verbs inflected for Wh-Agreement (*dumedengkut*) unambiguously signal the RC gap's position. We tested how participants resolved RCs using an auditory picture matching task, given on tablet computers. We instructed participants (n=135) to push a symbol to one of two pictures, depicting possible interactions between the two NP referents (ex. below). The form of the instruction crossed RC TYPE and VERB FORM (3×3 design; see Table). Each participant heard 36 items and we recorded their ultimate interpretation, RT and touch trajectory [cf. 6].

Results. Participants overwhelmingly interpreted transitive post-nominal RCs as SRs (94%). In contrast, the pre-nominal SR preference was only 43%. For passives, there were nearly no errors in the post-nominals; strikingly, in pre-nominals, the passive error rate was 30%. This pattern, compatible with a pre-nominal OR preference, is qualified by large inflections in the touch-trajectory toward the (correct) SR choice. In fact, the only reliable effect on trajectories was inflection toward SR pictures in trials when the OR picture was ultimately selected, regardless of RC type. Two further pieces of evidence bolster an underlying SR preference: (1) In headless relatives, the SR preference rose to 64%, suggesting the second NP plays a role in promoting the OR interpretation. (2) Substantial inter-island variation was observed in pre-nominal SR rates (ranging from 30% in Rota to 80% in Tinian), but not in post-nominal SR rates. This, we believe, is linked to inter-island variation in the integrity of the (Wh)-Agreement paradigm [7]. In speakers with limited subject Wh-Agreement, the SR preference receives relatively less competition from the agreement paradigm; and indeed subject Wh-Agreement error rates correlate with SR interpretations in pre-nominal but not post-nominal clauses.

RC TYPE (↓) × VERB FORM: {*trans:* HA DEDENGKUT | *wh.agr:* DUMEDENGKUT | *pass.:* DINEDENGKUT}		
	Chonnik i puti'un guatu gi atyu …	"Push the star over to that …"
POSTNOM	… na <u>sihik</u> [i **ha dedengkut** i gayu]	{ "… kingfisher that's pecking the rooster", "… kingfisher that the rooster's pecking" }
PRENOM	… [i **ha dedengkut** i gayu] na <u>sihik</u>	
HEADLESS	… [i **ha dedengkut** i gayu]	{"one pecking the rooster", "the r's pecking"}
NB: The RC head is underlined; the RC is bracketed; and the verb is in bold.		

References. [1] Keenan '77 *Linguistic Inquiry, 8.* [2] Vasishth '13 *PLOS One* [3] Clemens '14 *NLLT*
[4] Hsiao '03 *Cognition, 90.* [5] Yip, '07 *Studies SLA, 29.* [6] Freeman '10 *Behav Res Meth, 42.* [7] Wagers '12 *CUNY*

Resumptive pronouns and structural complexity in Cantonese relative clause production

Elaine J. Francis, Charles Lam, & Carol Chun Zheng (Purdue University), John Hitz (University of Central Oklahoma) & Stephen Matthews (University of Hong Kong)
ejfranci@purdue.edu

The use of resumptive pronouns (RPs) in relative clauses (RCs) appears to be governed by structural complexity in grammar and processing. RP distributions across languages follow the NP Accessibility Hierarchy (Keenan & Comrie 1977): if the grammar allows RPs in one position (e.g. direct object), it allows them in more deeply embedded positions (e.g. oblique object). Hawkins (2004) predicts a parallel effect in usage: when the grammar allows either RP or gap, RPs should be preferred more as structural complexity increases, to facilitate processing.

Cantonese optionally permits RPs in several contexts (Matthews & Yip 2011), providing a good test case for Hawkins' prediction. Two experiments examined three related questions: (1) Does optional RP use increase with structural complexity? (2) Are RPs preferred more in direct object than subject RCs, given the lack of such contrast in experiments on Mandarin RPs (Ning 2008; Yuan & Zhao 2005) and mixed results regarding subject-object asymmetry in the processing of Mandarin RCs with gaps (Chen et al 2008; Cheng et al 2011; Hsiao & Gibson 2003; Lin & Bever 2006)? (3) Are RPs required to prevent a syntactic island violation when the object of a coverb (preposition-like serial verb) is relativized, as claimed for Cantonese (Francis & Matthews 2006)?

Judgment Experiment: Participants listened to sentences twice and rated them on a seven-point scale. RP and gapped RCs were varied along two complexity-related dimensions: possession (non-possessive, possessive), and grammatical role (subject, direct object, coverb object). Five lexicalizations of each of the twelve conditions, along with equal numbers of filler sentences of varying degrees of acceptability, were included.

Production Experiment: Participants listened to two short sentences and combined them into one longer spoken sentence containing a subject-modifying RC. Target productions resembled acceptability stimuli, except that speakers' choice of RP or gap was a dependent variable.

The same 22 native-speaker participants completed both tasks. To ensure that participants were not aware of the RP manipulation in production, the production task was conducted first, followed by the acceptability task one week later. Results showed that RPs were possible in all positions, and were preferred over gaps in possessive RCs and non-possessive coverb RCs (Figures 1-2), supporting Hawkins' prediction that optionality is modulated by structural complexity. Contrary to studies of Mandarin, RPs were more acceptable (Figure 1) and occurred more often (Figure 2) in direct object than subject RCs. RPs were preferred but not required in coverb RCs, suggesting that coverb RCs might not be strict syntactic islands. Results were significant in mixed model regression analyses (all p's < .01).

Figure 1: Acceptability

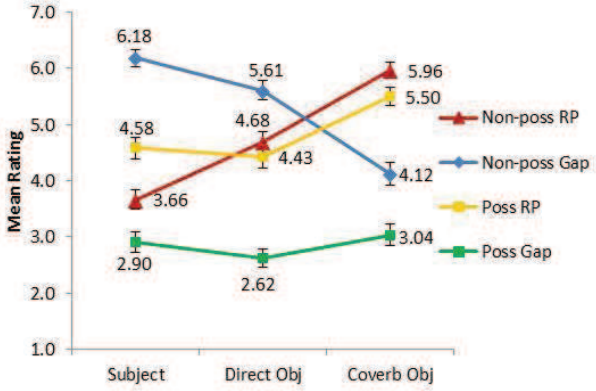
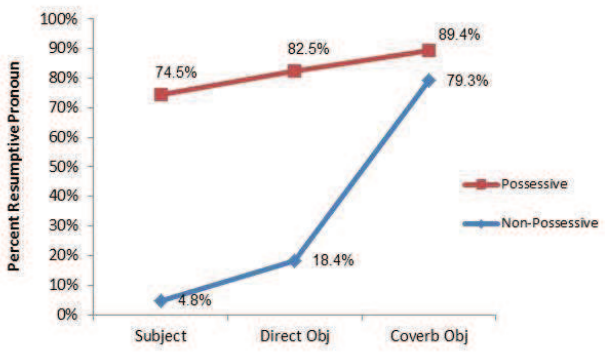


Figure 2: Production



Resumptive pronouns salvage island violations in forced-choice tasks

Lauren Ackerman, Michael Frazier, and Masaya Yoshida (Northwestern University)
lmackerman@u.northwestern.edu

Introduction: In the syntax literature, it is frequently reported that resumptive pronouns (RPs) can salvage or improve island violations (e.g., [1,2]). However, much psycholinguistics literature suggests RPs are no more acceptable than equivalent illicit gaps (see [3,4,5]; Cf. [6]). One explanation for the apparent discrepancy is the methodological differences between the two approaches. In psycholinguistics studies, RP and illicit gap conditions are not typically shown together. That is, participants do not directly compare the two conditions. In the syntax literature, the reported effects come from direct comparison of well-controlled minimal-pair sentences. Furthermore, in informal acceptability judgment, many native speakers detect acceptability differences between examples with an RP in an island and those with a gap in an island. Where does this discrepancy come from? There are several possibilities: formal acceptability experiments may not detect statistically significant salvation effects from RPs because these effects do not exist, i.e. RPs simply do not salvage island violations. Alternatively, salvation effects may be real but too weak for native speakers to detect unless a minimal pair is shown so that the alternatives (gap and RP) can be compared directly.

Experiment: We conducted a forced-choice sentence completion experiment. In this task, sentences are presented with a blank in the position of the relevant trace and the word preceding it as in (1), here marked with an underscore.

(1) a. Which man did Jane say that the parent who _____ forgave the babysitter's mistake?
scolded scolded him
b. Which man did Jane say that the parent who scolded the caregiver _____?
forgave forgave him

Participants indicated which of two options would best complete the sentence, either the preceding word and the RP, or the preceding word alone. In a 1x2 design the position of the gap/RP (within a relative clause island (RCI) vs within the matrix clause) was manipulated and participants' choice to complete the structure with a gap or RP structure was measured.

If RPs can salvage island violations, we expect the RP option will be selected more often than the gap option in the RCI condition (a). However, if RPs can*not* salvage islands, we expect that the two options will be chosen at chance level in RCIs, since both are equally unacceptable. In both scenarios, we expect the gap option to be preferred in the matrix clause condition (b), where the gap is licit.

The results were analyzed using a 1-sample proportions test with continuity correction. Our results show a strong preference for RPs in the RCI condition ($z=9.88$, $p<0.0001$: 82.1% RP, 17.9% Gap), but a preference for a gap in the matrix clause condition ($z=3.68$, $p<0.001$: 37.9% RP, 62.1% Gap). These results suggest the following. First, the fact that gaps are preferred to RPs in the matrix clause but not in the island suggests that the insertion of RPs is sensitive to grammatical structure: when there is no island to violate, RPs are dispreferred. Second, the fact that RPs are strongly preferred to gaps in islands suggests that RPs indeed improve island violations. Somewhat unexpectedly, there is a non-negligible number of RP completions even in the matrix condition (37.9%). This is not expected if RPs are not licit in non-islands ([4]). We speculate that this effect is due to the length and complexity of the target sentences, which has been suggested to be a source of increased acceptability for RPs [7].

Conclusion: We hypothesize that previous psycholinguistic work may have failed to detect the ameliorating effect of RPs on extraction out of islands because it did not require participants to compare. Our study indicates that when participants must choose between an RP continuation and an island-violating gap continuation, they display amelioration effects in accord with those reported in the syntax literature, consistent with observations (e.g. [3]) that more production-oriented tasks yield more evidence for the acceptability of RPs.

References: [1] Prince (1990). [2] Shlonsky (1992). [3] Heestand, Xiang & Polinsky (2011). [4] Alexopoulou & Keller (2007). [5] Omaki & Nakao (2004). [6] McDaniel & Cowart (1999). [7] Erteschik-Shir (1992).

Contrast-driven phonetic variation in English and Russian

Clara Cohen (University of California, Berkeley)
cpccohen@berkeley.edu

Probabilistic usage patterns in speech can be characterized in two broad ways: *contextual* probability and *paradigmatic* probability. Contextually probable units tend to be phonetically reduced [e.g., 3]. By contrast, paradigmatically probable units — typically morphemes in words with a high relative frequency within a morphological paradigm — show both enhancement [4] and reduction [3]. This pronunciation variation affects different phonetic features, including duration, vowel centralization, and segment deletion. Here, I ask (a) how these features are determined, and (b) why effects of paradigmatic probability are inconsistent. The results of the following two experiments suggest that the structure of a language's inflectional paradigm influences which phonetic features show pronunciation variation.

In Experiment 1, 40 English speakers produced 60 sentences with complex subject NPs (e.g., *The key$_{SG}$ to the cabinets$_{PL}$*), and singular verbs ending in the suffix {-s} (e.g., look-s). Such complex subjects allow both singular and plural verb agreement at well-studied rates [e.g., 1]. This design therefore allowed the {-s} to vary in contextual probability. The 52 verbs used in the experiment had relative frequencies within their paradigms; as a result, the suffix also varied in paradigmatic probability. Log-transformed suffix durations were analyzed with mixed effects regression modeling,with random intercepts for subject and sentence, and fixed effects for contextual and paradigmatic probability. Other fixed effects controlled for speech rate, pauses, stem length, phonotactics, and verb frequency. The log-transformed ratio of suffix-to-stem duration was analyzed with a similar model. Relative suffix duration showed an interaction between contextual probability and verb frequency: Compared to stem duration, suffixes were relatively shorter as their contextual probability increased, but the effect was reversed for high-frequency verbs. Absolute suffix duration was lengthened as paradigmatic probability increased, thus showing enhancement, parallel to [4].

In Experiment 2, 44 Russian speakers produced 76 sentences with quantified subject noun phrases. Such sentences allow either singular verb agreement, (marked here by {-o}), or plural agreement (marked with {-i}). The probability of plural agreement in these sentences varied between 0.01 and 0.99, as calculated by a norming study. Each sentence contained one of 35 end-stressed verbs with either singular or plural agreement, counterbalanced by list. Suffix duration, F1, and F2 were analyzed with mixed effects regression modeling, parallel to Experiment 1. Duration and F1 showed no effect of either contextual or paradigmatic probability. By contrast, increasing paradigmatic probability yielded significant phonetic enhancement of F2: F2 decreased for {-o} (backing of a back vowel) and increased for {-i} (fronting of a front vowel).

Together, both experiments suggest an explanation for the previously observed inconsistent effects of paradigmatic probability on pronunciation. The phonetic feature that varies with paradigmatic probability depends on the nature of the contextually relevant morphological contrast. In both experiments, the distinction at issue was between singular and plural verbs. In English, this distinction is encoded by an [s]/ ∅ contrast, and in English the enhancement effect of higher paradigmatic probability appeared on suffix duration. In Russian, the distinction was encoded by vowels that differed most saliently in backness, and the enhancement effect appeared on F2. Higher paradigmatic probability does therefore lead to phonetic enhancement, but the feature that is enhanced depends on which morphological contrast is relevant in the utterance, and what phonological feature marks that contrast.

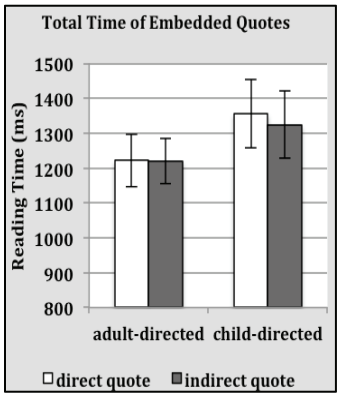
[1] Eberhard, K. M., J. C. Cutting, & J. K. Bock. 2005. Psychological Review 112:531 – 559. [2] Hanique, I., and M. Ernestus. 2011. Interspeech 2011, 2849-2852. [3] Jurafsky, D. , A.Bell, M. Gregory, & W. Raymond. 2001. In Frequency and the emergence of linguistic structure, ed. J. Bybee & P. Hopper. Amsterdam: John Benjamins. 229 – 255. [4] Kuperman, V., M. Pluymaekers, M. Ernestus, & R. H. Baayen. 2007. JASA 121: 2261 – 2271.

Auditory perceptual simulation of child-directed speech in silent reading

Zhiying Qian, Mallory Stites & Kiel Christianson (University of Illinois at Urbana-Champaign)
zqian3@illinois.edu

Previous studies show that readers perceptually simulate aspects of speakers' speech during silent reading. Yao & Scheepers (2011) observed shorter reading times for direct quotes said by a fast talker compared to a slow talker. Stites et al. (2013) similarly observed faster reading of direct quotes when the adverb describing speech rate as fast than when described as slow (e.g. "*John walked into the room and said quickly/slowly, "I finally found my car keys*"). Most recently, Zhou and Christianson (2013) report that readers also perceptually simulate non-native speech after viewing a picture or hearing the name of non-native and native speakers. These studies suggest that perceptual simulation can be triggered by explicitly describing the speech rate or the talker's characteristics (e.g. fast vs. slow speakers; native vs. non-native speakers), but leave open the question of whether readers can engage in auditory perceptual simulation (APS) when these cues are absent. In the present study, we manipulated the characteristics of the depicted listener rather than the depicted speaker, specifically, whether the listener was a child or an adult, to determine if readers simulate child-directed speech in the absence of information about the speaker or speech rate. In addition, we tested whether APS effects could be observed in a non-alphabetic orthography, where grapheme-phoneme mappings are notoriously opaque.

Native-speaking Chinese participants (n=31) read Chinese sentences from one of four lists containing 28 experimental items and 80 fillers while their eye movements were monitored. Experimental items followed a 2 (adultness: child-directed vs. adult-directed speech) x 2 (directness: direct vs. indirect speech) design. Experimental items shared the same template of "somebody said to a child/an adult, followed by a direct/indirect quote" as in: *I said to the little boy/the young man who liked drawing, (")the bridge is very well drawn(")*. The stimuli observed Chinese conventions of punctuation for direct quotes, i.e. direct quotes were preceded by a colon and a quotation mark, and for indirect quotes, which were preceded by a comma. Reading time measures were obtained for the regions of the embedded direct and indirect quotes *(*i.e., *the bridge is very well drawn*).



LME modeling revealed a significant main effect of adultness ($p<0.05$) (M difference between child- vs. adult-directed speech = 119ms); the same pattern held for go-past time ($p=0.05$, M difference = 156ms), showing that readers spent more time reading speech directed at children than adults. Unlike previous results in alphabetic orthographies, this APS effect was present regardless of whether the speech was a direct quote or reported speech.

Our results extended previous findings by demonstrating that readers can use cues other than explicit descriptions of the speaker or speech rate to guide perceptual simulation. Also, the results suggest that, at least in reading Chinese text, APS of child-directed speech was performed for both direct and indirect quotes. There are two possible explanations for this effect: 1) It is possible that simulation of speech rate in indirect quotes is specific to tonal languages, as the comma in indirect speech may not be a strong enough cue to suppress the lengthening of lexical tones in child-directed speech; and 2) The comma in indirect quote condition (e.g. *I said to the little boy who liked drawing, the bridge was very well drawn*) segmented the speech from its preceding text in a format resembling the direct speech condition (*I said to the little boy who liked drawing: "The bridge is very well drawn"*). Thus readers may have processed the indirect quote condition similarly to the direct quote condition. Further study is on-going to tease these two factors apart.

Listeners update probabilistic percepts of (a) determiner(s) several syllables downstream

Meredith Brown (University of Rochester; mbrown@bcs.rochester.edu), Laura C. Dilley (Michigan State University), & Michael K. Tanenhaus (University of Rochester)

It is commonly assumed that spoken language processing is radically incremental and involves rapid abstraction from acoustic-phonetic detail to categorical representations. An alternative view holds that listeners maintain and integrate probabilistic perceptual information over a longer temporal window (Levy et al., 2009; Dahan, 2010; Johnstone et al., 2013). We present results from two visual-world experiments indicating that listeners use speech rate to generate provisional hypotheses about the timing and number of constituents in portions of an utterance, and update these provisional hypotheses based on downstream prosodic information.

Experiment 1 used 54 utterances containing polysyllabic nouns preceded by heavily-coarticulated indefinite articles and followed by sibilant-initial words (Fig. 1, top). We altered the duration of the determiner "a" and the segments surrounding it (*determiner region*) to manipulate the likelihood of its perception (Dilley and Pitt, 2010). We also altered the speech rate of material distal to the critical sibilant (*sibilant region*) to manipulate the likelihood that the target word would be perceived as ending in the plural morpheme *–s* (Fig. 1, bottom). Participants (*n*=36) were asked to select the mentioned referent within a four-picture display containing pictures of singular and plural versions of the target noun (e.g. one vs. two raccoons) and of an unmentioned noun while their eye-movements were monitored. Proportions of plural vs. singular responses were significantly affected by the speech rate of both the determiner ($p<.0001$) and the sibilant relative to surrounding context ($p<.0001$): Higher proportions of plural responses were associated with determiners perceived to be short and with sibilants perceived to be long. Likewise, eye-movements revealed rapid effects of determiner duration: Shorter determiners elicited higher proportions of plural target fixations than longer determiners prior to ($p=.052$) and after the processing of the sibilant ($p<.0005$), demonstrating that listeners used acoustic cues to the presence or absence of the determiner shortly after they became available. Importantly, they also made rapid use of the sibilant's duration relative to context speech rate, as indexed by proportions of fixations to plural vs. singular target pictures following the onset of the sibilant ($p<.0005$). These findings indicate that determiner perception is influenced by prosodic information multiple syllables downstream.

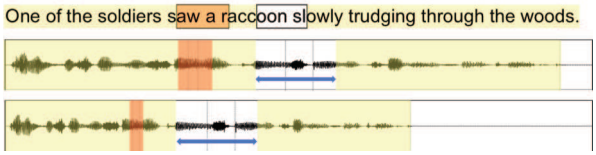


Fig. 1. Example stimulus (top) and illustration of the distal speech rate manipulation (bottom).

Experiment 2 addressed the possibility that the apparent effects of sibilant speech rate relative to surrounding context in Experiment 1 could instead be attributed to effects of distal speech rate compression on the absolute duration of the determiner (cf. Fig. 1). Its design was identical to Experiment 1, except that we manipulated the speech rate of the sibilant region directly, rather than manipulating the speech rate of surrounding context. This ensured that the absolute duration of the determiner did not covary with the speech rate of the sibilant. Results from 36 participants indicated that the speech rate of the sibilant continued to influence proportions of plural responses ($p<.0001$) and fixations to plural target pictures following sibilant onset ($p<.0005$), providing stronger evidence that the interpretation of function words can be modulated by information encountered considerably later in the utterance.

Taken together, these results suggest that listeners maintain and update provisional hypotheses about previously encountered material across multiple morphophonemic units. These findings are most naturally explained within a probabilistic data-explanation account of spoken language processing: Probabilistic inferences about the hidden causes underlying aspects of an utterance (such as the intention to produce a singular or plural construction) inform fine-grained probabilistic expectations about how aspects of lexical alternatives will be realized in context, and can be updated in light of subsequent cues in the unfolding utterance.

Great expectation follows big surprise: Direct evidence for syntactic adaptation from eye-movements

Manabu Arai (The University of Tokyo, RIKEN BSI) & T. Florian Jaeger (University of Rochester)
m-arai@phiz.c.u-tokyo.ac.jp

Recent exposure to a particular syntactic structure changes the likelihood of the same structure being produced or expected in a subsequent utterance. Error-based accounts hold that the prediction error experienced while processing a prime leads to adapted expectations/use in subsequent trials (Chang et al., 2006; Jaeger & Snider, 2013). This in turn has been claimed to allow comprehenders to converge towards to context-specific statistics of syntactic structures, thereby lowering the average surprisal experienced (Fine et al., 2013; Kamide, 2012). Evidence for error-based syntactic priming in comprehension comes from Fine and Jaeger (2013). What is lacking so far is a) behavioral evidence for expectations while processing a prime and b) a direct link between that prediction error and the magnitude of syntactic priming.

Data: We re-analyze the eye-tracking data of adult participants in Arai and Mazuka (2014), who found a robust (lexically independent) priming effect in the comprehension of Japanese passive structure. The measure of the priming effect was participants' anticipatory looks toward either agent or patient entities in a visual scene upon hearing a case-ambiguous sentence-initial NP and an adverb in target sentences following an experience of either an active or passive prime. The order of individual experimental items was randomized, eliminating confounding of item and trial order.

Prediction 1: Changing expectations during prime processing – Since passives are less frequent than actives, passives should be unexpected at the beginning of the experiment. If comprehenders integrate prior knowledge with the observations made in the experiment, they should come to increasingly expect passives as the experiment progresses. As a measure of participants' expectations for passives, we analyzed the proportion of fixations toward the patient (the first mentioned entity in passive primes) during the 0-500 ms interval from the prime's speech onset. We found the predicted pattern: The probability of fixations on the patient for passive primes was lowest at the beginning of the experiment and gradually increased across trials. The probability during active primes, in contrast, did not change throughout. All analyses used mixed logit models with full random effect structure. All reported effects are at $p<.05$.

Prediction 2: Changing magnitude of prime effect – If error-based syntactic priming underlies these changes in expectations through the course of the experiment, this should also be reflected in changes in the *magnitude* of priming (the amount of shift in expectations during processing of the *target* structure). As participants' estimates of the passive vs. active distribution converges towards equi-probability (due to the balanced design), the average prediction error experienced for active and passive primes should decrease (this follows from the observation that the prediction error assumed in error-based models, Chang et al., 2006, is log-linear in probability, Jaeger and Snider, 2013). We analyzed the ratio of fixations to the agent and patient as a measure of participants' expectations. We replicated the prime effects reported in Arai and Mazuka (2014) and found that it decreased linearly through the experiment.

Conclusions: The results together demonstrated that the magnitude of surprise on hearing the passive structure in prime trials was directly reflected in the magnitude of the expectation about the passive structure in target trials; The greater the prediction error in a prime trial, the more strongly participants predicted the passive structure in a subsequent target trial. The results provide the most direct evidence to date for dynamic adaptation of grammatical knowledge to current linguistic input.

Reference

Arai, M. & Mazuka, R. The development of Japanese passive syntax as indexed by structural priming in comprehension. *The Quarterly Journal of Experimental Psychology. 67*, 60-78, 2014.

Contributions of hippocampal-dependent declarative memory to on-line processing of global syntactic ambiguity

Sarah Brown-Schmidt (University of Illinois), Zhenghan Qi (MIT), & Melissa Duff (University of Iowa)
sarahbrownschmidt@gmail.com

An increasing consensus is that the hippocampal-dependent declarative memory system contributes to a variety of cognitive domains beyond its traditional role in declarative memory. We propose the hippocampus, in its capacity for relational binding, representational flexibility, and on-line representational processing (Eichenbaum & Cohen, 2001; Hannula, et al., 2006), is a key contributor to language processing. Here we examine hippocampal contributions to the use of distributional cues to verb argument structure in on-line language processing. We test a novel claim, that if these cues can be used independently of the declarative memory system, then this process should be inflexible. We test this hypothesis by examining performance of hippocampal amnesic patients with severe declarative memory impairment in an ambiguity resolution task. Syntactic ambiguities such as *Feel the pig with the flower* have two interpretations: In one, the prepositional phrase (PP) modifies the noun, *pig*; in the other the PP specifies an instrument. Healthy adults are sensitive to differences across verbs in the likelihood that a PP will modify the noun or the verb (Snedeker & Trueswell, 2003; 2004). Whether an intact declarative memory system is necessary to use these verb biases during processing is unknown. We hypothesized that (1) verb bias may be used independently of the declarative memory system, but that (2) in such cases, processing is inflexible.

METHOD: Participants were 4 amnesic patients with bilateral hippocampal damage and severe declarative memory impairment, and 4 healthy matched comparison participants. Patients were free of aphasia. Eye movements were tracked as participants followed instructions. Critical scenes contained 2 animals, e.g., a frog holding a small flower and a zebra holding a small candle, and two larger items, a flower and a candle. The critical manipulation was whether the verb was biased to take a PP instrument argument (e.g., *Hit the frog with the flower*), equi-biased (e.g., *Feel*) or modifier-biased (*Find*). Actions (whether the participant attached the ambiguous PP to the verb, and therefore used an instrument to carry out the action, or not), and eye-gaze to the potential instrument (flower) and target animal (pig) were analyzed with mixed models; results are significant at $p<.05$ unless noted.

RESULTS: A control study with 13 healthy undergraduates replicated Snedeker & Trueswell (2004): participants acted out instrument interpretations for instrument-bias (I) verbs more than modifier-bias (M) verbs, with equi-bias (E) verbs in-between.

Amnesic patients and healthy comparison participants showed a verb-bias effect in their actions; in fact, the difference between (I) and (E) verbs was larger for patients. This shows that amnesic patients effectively used verb bias to guide their final interpretation. However, analysis of gaze revealed (1) Patients were slower to use verb bias: During interpretation of *Hit the frog*, comparisons, but not patients showed the expected verb-bias effect in looks to the animals (more animal looks for (M) verbs). Also, during *flower*, patients looked more at the incorrect animal, particularly when the additional verb argument was introduced in the case of (I) verbs. (2) Patients were less flexible in their use of verb bias: During *flower, c*omparisons, but not patients, processed (I) verbs flexibly on-line, with more target animal ($p=.06$) and fewer instrument fixations when acting out modifier vs. instrument interpretations of (I) verbs.

SUMMARY: Amnesic individuals with severe declarative memory impairment did use verb bias to guide the ultimate syntactic interpretation of sentences. However, during on-line processing, these verb bias effects emerged more slowly, and were less flexible. These findings provide key evidence for hippocampal contributions to on-line language processing. The mechanism of this contribution may be the rapid and flexible on-line integration of distributional cues with the unfolding sentence.

Morphology aids syntax in noisy sentence processing

Gaurav Kharkwal & Karin Stromswold (Rutgers University – New Brunswick)
kharkwal@rutgers.edu

Noisy-channel models of sentence processing suggest input uncertainty is maintained and subsequently resolved using prior grammatical information (Levy, 2008, 2011). Arguably, the first step in sentence processing is determining the syntactic categories of words. We explored the extent to which syntactic and morphological information are used in sentence processing to determine the syntactic categories of unknown words (pseudowords).

Experiment 1. Thirty-six native, monolingual English speakers read 60 target sentences (subject and object-extracted center embedded relative clauses) and 120 filler sentences in a non-cumulative word-by-word format. In half the sentences, a single lexical word at position 2, 3, or 4 was replaced with an orthographically and phonologically plausible pseudoverb with an *-ed* suffix (*scoaned*) or a morphologically unmarked pseudonoun (*threak*).

[1] **S**ubject **RC**: *The actor$_1$ who impressed$_2$ the critic$_3$ humiliated$_4$ the director$_5$.*

[2] **O**bject **RC**: *The actor$_1$ who the critic$_2$ impressed$_3$ humiliated$_4$ the director$_5$.*

Results. Per-word analyses were performed on residual RTs (RT - word length) using mixed-effects models with random intercepts for Subject and Item. Consistent with earlier work (Grodner & Gibson, 2005), for non-pseudoword sentences, participants were slower overall for ORCs, with the RC verb being the point of major difficulty (p<.05). Analyses comparing pseudoword vs. non-pseudoword sentences revealed that reading pseudowords did not always incur a penalty. For SRCs, only a pseudoword at position 2 caused a significant slowdown, which persisted through the end of the RC clause (p's < .05). For ORCs, pseudowords at both positions 2 and 3 caused a significant slowdown, but the slowdown did not persist (p's < .05).

The lack of a consistent pseudoword processing penalty suggests, at the very least, that determining the syntactic category of unknown words is seamless, and possibly reliant on the syntactic context. Next, we investigated whether morphological information plays a role as well.

Experiment 2. Thirty-two native, monolingual English speakers participated. The experiment was the same as in Experiment 1 with the following exceptions. First, we only substituted pseudowords at positions 2 and 3. Second, pseudowords were morphologically "congruent" with their syntactic position half the time, and "incongruent" in the other half. Congruent pseudonouns had a nominal derivational suffix (*-or, -ist, -ician*) and a plural suffix (*-s*), and congruent pseudoverbs had a verbal derivational suffix (*-ify,-ize, -ate*) and a past-tense inflectional suffix (*-ed*). Verbal and nominal suffixes were interchanged in non-congruent words.

Results. As in Experiment 1, subjects were slower overall for ORCs, with the point of major difficulty being the RC verb (p<.05). Unlike in Experiment 1, for both SRCs and ORCs, participants slowed down at both pseudoword positions (p's <.05). At the actual pseudoword positions, no significant differences were found between morphologically congruent vs. incongruent pseudowords. However, difficulty in processing incongruent morphology was observed in the form of longer RT "spillover" for morphologically incongruent pseudowords. For SRCs, slowdown persisted for two words after incongruent pseudowords, but only one word for congruent pseudowords. In addition, for pseudowords at position 2 of SRCs (the only syntactically ambiguous position), participants were significantly slower at the following word when the pseudoword was morphologically incongruent than when it was congruent (p<.05). For ORCs, the slowdown only persisted through the following word for incongruent pseudowords (p's<.05), and did not persist at all for congruent pseudowords.

Conclusion. Consistent with noisy channel models, our findings indicate an important role of syntactic context in determining syntactic categories of unknown words. Crucially, when syntactic information is inadequate (e.g., position 2 in [1]), morphological information seems to guide sentence processing. Overall, our findings highlight the robustness of sentence processing and the interplay of morphological and syntactic information in guiding it.

Working-memory capacity modulates antilocality effects in syntactic dependencies

Bruno Nicenboim & Shravan Vasishth & Reinhold Kliegl (University of Potsdam) & Carolina Gattei (Institute of Biology and Experimental Medicine) and Mariano Sigman (Laboratory of Integrative Neuroscience)
bruno.nicenboim@uni-potsdam.de

Increasing dependent-head distance has been claimed to lead to more processing efforts at the head (*had fired* in 1) of the dependency (locality: Gibson, 2000; Lewis & Vasishth, 2005). However, recent work has found the opposite effect (anti-locality; for examples, see Levy 2008): facilitation with increasing distance. Levy explains this effect as facilitation due to stronger expectations for the upcoming VP. Although it is clear that working memory capacity (WMC) is involved in locality effects, expectations are generally treated as orthogonal to WMC accounts (e.g., Levy 2008). We show that WMC determines whether expectation-driven facilitation is possible.

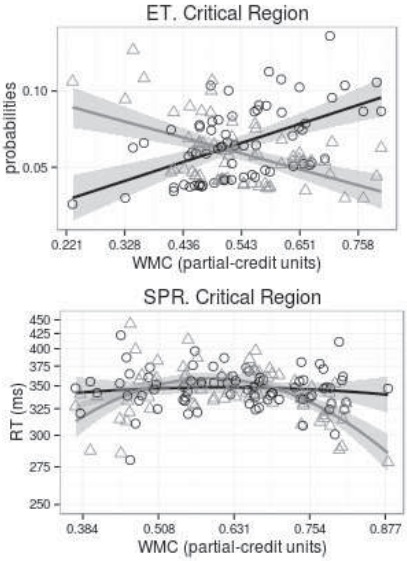
We conducted two experiments in Spanish (eyetracking, N=71; SPR, N=79) using sentences such as (1); for space reasons a subset of the conditions and results are discussed. The distance between the *wh*-element and the head was manipulated by interposing a AdvP in (b). The baseline condition (a) had a shorter dependency, but the verb was kept in the same position of the sentence. Participants' WMC was measured using an operation span task (Turner & Engle, 1989). Partial-credit was used as a score of individual WMC (Conway et al., 2005).

1. a. [$_{AdvP}$ *before some days*,] José asked **who.ACC [$_{S(interm)}$** they-said that **[$_{S}$** the manager **had fired** because …**]]**

b. José asked **who.ACC [$_{S(interm)}$** [$_{AdvP}$ *before some days*,] they-said that **[$_{S}$** the manager **had fired** because …**]]**

Predictions. Locality predicts more difficulty when distance is increased, whereas expectation predicts facilitation. If WMC plays a role, locality effects should appear only for low-capacity readers; while an expectation-driven facilitation should appear when WMC is large enough so that readers can overcome the memory load produced by the increased distance and maintain predictions of upcoming material.

Results. In the eye tracking study, we found an interaction between WMC and distance for first pass regression probabilities, z=-2.2: as WMC increased, probabilites of a regression were lower in (b) than in (a). I.e., locality for low-WMC readers and an expectation-based facilitation for higher WMC. In the SPR study, at the critical verb region, we found an interaction between WMC² and distance, t=-2.8. This was an inverted U-shaped effect of WMC on RT, that is, shorter RTs in (b) vs (a) for low and high-WMC than for mid-WMC. I.e., speedups were seen in low as well as high-capacity readers, but not in medium-capacity readers.



(a) black circles; (b) grey triangles

Conclusions. The eye tracking study showed that expectation-driven facilitation depends on having enough capacity to overcome the memory load produced by increased distance and perhaps also on a better ability to maintain predictions. The SPR data showed a very surprising inverted U-shaped WMC x expectation interaction. One possible explanation is that when low-WMC readers parse sentences while engaged in memory taxing processes, they may disengage from the memory load sooner by reading *faster* (using a good-enough parsing heuristic, see Ferreira & Patson, 2007), especially due to the increased task demands of SPR. By contrast, high-WMC readers could be showing expectation-based facilitation because they are able to overcome locality effects and maybe also maintain predictions better. In sum, this is the first study showing that WMC modulates anti-locality effects.

Efficient Communication Forwards and Backwards

Richard Futrell (MIT), Kyle Mahowald (MIT), Steve Piantadosi (University of Rochester), & Edward Gibson (MIT)
Contact: *futrell@mit.edu*

Information-theoretic studies of human language start by quantifying the informativity of linguistic material. This requires measuring how predictable the material is in context. Communicative efficiency demands that more informative material be more prominent and that less informative material be reduced, whether in length or articulatory effort (Shannon, 1948). But if informativity is predictability in context, that raises the question: what context? We present three results showing that word length varies according to predictability given *following* context; this finding limits the set of audience design theories compatible with the data.

Most studies along these lines examine the effects of predictability given preceding words. For example, Piantadosi et al. (2011) correlate word lengths with surprisal given the preceding two words, and Jaeger (2010) predicts "that" omission using the probability of a complement clause given the preceding verb. Even Shannon's (1951) estimate of the entropy of English uses preceding context only. We call surprisal estimated this way "forward surprisal."

Here we study "backward surprisal": the predictability of words given *following* words. A previous study, Bell et al. (2009), finds that word duration in English speech is a function of probability given following material, and *not* of probability given preceding material.

First, in *N*-gram corpora for 13 languages, we replicate Piantadosi et al.'s (2011) finding that orthographic word lengths are better correlated with surprisal than frequency. The correlation of length with backward surprisal is just as strong or stronger than the correlation with forward surprisal (Figure 1). Second, we replicate Mahowald et al.'s (2013) finding that shorter word variants appear in more predictive contexts in corpora; we find that backward surprisal also predicts the use of short variants. Third, in a behavioral study, we find that subjects choose the shorter word variants given informative context both before and after the critical word.

The effect of backward surprisal is not compatible with a certain subset of audience design theories. Because the speaker is reducing his signal on the basis of context the hearer does not know, the information theoretic reduction effects cannot be the result of the speaker trying to prevent momentary spikes in surprisal for the hearer. It is compatible with theories where reduction is driven by efficiency for the speaker, or by theories where the speaker is trying to control the local but not immediate information density of the signal for the hearer.

The relevance of preceding context for surprisal seems intuitively obvious: it is the order in which linguistic units are sent and received. But in the context of the whole processes of production and comprehension, it is less clear that forward context is what matters. In producing an utterance, a speaker has an idea what she is going to say. Similarly, if one part of the utterance is surprising when the hearer initially hears it, that is fine if following parts of the utterance make it make sense in retrospect. We hope the demonstrated effects of backward surprisal motivate a move away from surprisal models that use only preceding context.

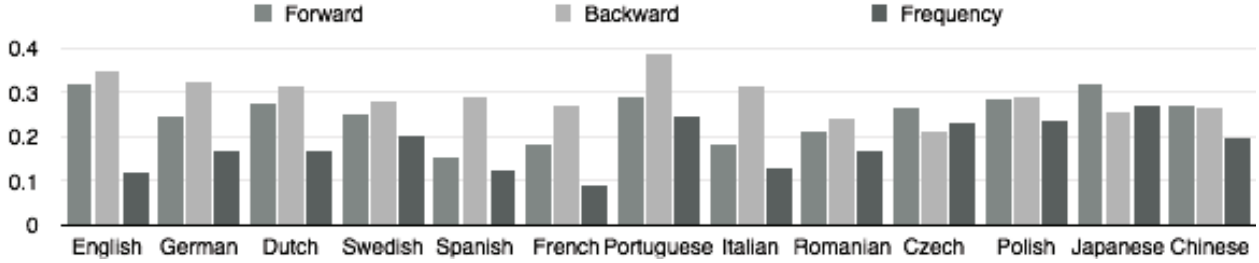


Figure 1. Spearman correlation coefficients of word length with forward surprisal, backward surprisal, and frequency in 13 languages.

Partial use of available information in the early stages of verb prediction

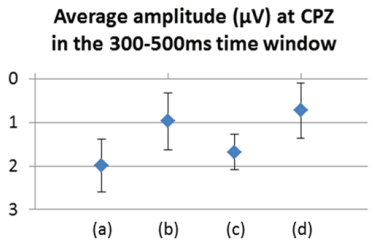
Wing Yee Chow (Basque Center on Cognition, Brain and Language), Glynis MacMillan, Shefali Shah, Ilia Kurenkov, Ellen Lau & Colin Phillips (University of Maryland)
wingyeechow.zoey@gmail.com

Previous research has shown that comprehenders use contextual information to anticipate upcoming input on the fly [1-2]. However, less is known about how different sources of information contribute to on-line predictive computations. Recent event-related potential (ERP) evidence indicates that early stages of verb prediction are based on somewhat superficial information. The predictions are affected by what words precede the verb in the sentence, but surprisingly they are not affected by the thematic roles of the words in the sentence [3]. This is demonstrated by the finding that the N400, a negative-going ERP component associated with lexical semantic processing [4], is differentially sensitive to a verb's off-line predictability (*cloze probability*) depending on how the context was manipulated. In particular, in verb-final clauses, the N400 was modulated when cloze probability of a verb was manipulated by substituting one of its arguments (e.g., "*The secretary confirmed which illustrator/readers the author had hired…*"), but not when a cloze probability manipulation of the same size was achieved by simply reversing the arguments (e.g., "*The restaurant owner forgot which customer/waitress the waitress/customer had served …*"). The current study asks: just how superficial are early verb predictions? Are they based on the meaning of all prior words in the context (in a "bag of words" fashion)? We argue that early verb predictions involve a more targeted mechanism, which specifically relies on the arguments in a clause to anticipate likely verbs, while initially failing to take into consideration the information provided by the thematic roles of those arguments.

In order to test which information impacts initial verb predictions, we varied the words that preceded a verb in English in two ways. The "different words" manipulation (*a* vs. *b*) involved the substitution of one of the verb's arguments only (e.g., "which exterminator" became "which neighbor"). The "same words" manipulation (*c* vs. *d*) involved the same argument substitution, but the words were further rearranged to create sentence contexts that contained an identical set of words. Both manipulations yielded an identical difference in the target verb's cloze probability (expected: 22% vs. unexpected: <1%). If initial verb prediction is based on a bag-of-words mechanism, which does not care about clause membership, then the "different words" manipulation should elicit an N400 effect, but the "same words" manipulation should not. Alternatively, if initial verb prediction is based only on the words in the same clause as the verb, then both manipulations should elicit an identical N400 effect. Results (n=23) revealed a significant main effect of cloze probability in the 300-500 ms interval, showing that the N400 was reduced for expected verbs compared to unexpected verbs, regardless of whether the sentence contexts as a whole contained the same set of words. (The N400 effect was additionally followed by a posterior late positivity (P600) in the "different words" conditions only, perhaps due to a slightly more sustained N400 effect in the "same words" conditions.) Critically, the preserved N400 sensitivity to the "same words" manipulation in the current study suggests that comprehenders can quickly target the arguments in a clause to predict an upcoming verb.

"Different words" argument substitution
(a) The tenant inquired [which *exterminator* the landlord had **hired**]…
(b) The tenant inquired [which *neighbor* the landlord had **hired**]…

"Same words" argument substitution
(c) The neighbor inquired [which *exterminator* the landlord had **hired**]…
(d) The exterminator inquired [which *neighbor* the landlord had **hired**]…



References: [1] DeLong et al. (2005) *Nat Neurosci.*, 8, 1117-1121. [2] Altmann & Kamide. (1999). *Cognition*, 73, 247–264. [3] Chow et al. (2013) *CUNY*. [4] Kutas & Hillyard. (1984). *Nature*, 307, 161-163.

Beyond syntax: Effects of verb semantics and perspective taking on Chinese reflexives

Xiao He & Elsi Kaiser (University of Southern California)
hex@usc.edu

We investigated how perspective-taking and verb semantics interact with binding-theoretic constraints during the processing of the Chinese reflexive *ziji* "self". According to Huang/Liu'01, Chinese referents' accessibility depends on intervening referents' person features: An intervening 3rd-person referent allows *ziji* to refer to local or long-distance (LD) subject (ex.1a&c); but an intervening 2nd-person referent blocks LD binding (ex.1b).

We used visual-world eye-tracking to **(i)** test when/how intervening subjects' person-features affect LD referents' accessibility in real-time. Furthermore, given debates about whether verb semantics are integrated predictively or only upon encountering an anaphor (e.g., **J**ärvikivi/Pyykkönen'10; see Schumacher et al.'11 on effects of verb bias on blocking in Chinese), we **(ii)** examined how/when verb biases modulate blocking. Furthermore, given claims from theoretical linguistics that Blocking involves *perspective-taking* (Huang/Liu'01), **(iii)** we built on insights from Brunyé et al'09 to probe for real-time evidence of perspectival effects.

(1a). **ni$_1$** shuo **Bill$_2$** zai gongyuan {**fanxing/jianshi**-le} **ziji$_{1,2}$**
you$_2$ say **Bill$_2$** at park {**reflect-upon/spy-on**-perf} **SELF$_{1,2}$**

(1b). ~~**John$_1$**~~ shuo **ni$_2$** zai gongyuan {**fanxing/jianshi**-le} **ziji$_{*1,2}$** ←BLOCKING
~~**John$_1$**~~ say **you$_2$** at park {**reflect-upon/spy-on**-perf} **SELF$_{*1,2}$**

(1c). **John$_1$** shuo **Bill$_2$** zai gongyuan {**fanxing/jianshi**-le} **ziji$_{1,2}$**
John$_1$ say **Bill$_2$** at park {**reflect-upon/spy-on**-perf} **SELF$_{1,2}$**

We manipulated **(i)** the form of the LD subject and the local subject (names/'you'), and **(ii)** verb bias (self-vs.-other-directed verbs, normed for Chinese by Schumacher et al.'11, cf. English *shave/kick*), as in ex(1). Participants listened to sentences while looking at corresponding pictures, and clicked on the last-mentioned thing (on targets, the referent of *ziji*). In **Exp.1,** 18 people were instructed to interpret the 2nd-person pronoun *ni* "you" as referring to some other person, in an 'overhearing' context (i.e., the participant is not the addressee). In **Exp2**, 18 new participants saw the same items, but now they were instructed to interpret the 2nd-person pronoun as referring to themselves (i.e., participants as addressees). This allows us to test whether (i) 2nd person blocking occurs regardless of the interpretation of the 2nd person pronoun, or whether (ii) perspective taking (participant=addressee) strengthens blocking.

Predictions: If Blocking determines antecedent choices, 2nd person blockers should render LD antecedents *unavailable* regardless of verb. But if verb biases modulate Blocking, listeners should look at LD subjects more with other-directed than self-directed. Furthermore, if verb biases cause listeners to activate more probable referents, people should look at verb-bias-consistent subjects early, perhaps even before the reflexive. In addition, if blocking involves perspective-taking, we expect more blocking in Exp.2 (addressee) than Exp.1 (overhearer).

Eye-movements show that blocking is not absolute and is modulated by verb biases and the addressee/overhearer distinction. Specifically, self-directed verbs cause an early preference for local subjects in both Exp1 and Exp2, and other-directed verbs trigger more looks to LD subjects, indicating that verb directedness kicks in early. In fact, even in the blocking condition, where LD subjects should be *un*available, other-directed verbs led to more looks to LD subjects (relative to local subjects) as early as 200*ms* after the onset of *ziji* in Exp1 (overhearer). However, in the same condition in Exp2 (addressee), more looks to LD subjects emerged much later – at 800*ms* after *ziji* onset. *Thus, blocking effects are stronger in Exp2, where 'you' refers to the participant.* Namely, although the blocking conditions in both Exp1 and Exp2 contain 'you', the 2nd person pronoun is a stronger blocker when it refers to the participant. **In sum**, our findings highlight the importance of non-syntactic factors (verb semantics and perspective-taking) on the real-time processing of reflexive anaphors.

2.5-year-olds consider presuppositions: An eye-tracking study on TOO and AGAIN

Frauke Berger (University of Potsdam) & Nausicaa Pouscoulous (UCL)
fberger@uni-potsdam.de

Sentences containing the additive expressions *too* or *again* presuppose that a contextually salient event has taken place in addition to the one asserted. For instance*, Lissy danced, too* presupposes another person has been dancing, while *Lissy danced again* presupposes the same person danced before. The meaning contribution of *too* or *again* thus strictly relate to background information. Results of previous studies using sentence picture matching tasks or picture selection tasks indicate that children up to school age do not consider additive particles in sentence interpretation (e.g., Bergsma, 2006), despite producing the particle from early on (around their 2nd birthday) (e.g., Nederstigt, 2003). We argue the apparent delay in comprehension to be task-related: Demonstrating the target interpretation in such tasks required (a) to accommodate the presupposition, (b) to detect its non-fulfillment in the visual context, and (c) to reject the sentence (or picture) for this reason. Taking results from adult studies into account (e.g., Schwarz, 2007; Kim, 2008; Cummins et al., 2013), all three task demands might be argued to reduce or mask children's potential target interpretation of sentences containing *too* or *again*. In addition, when using an alternative experimental task, 3-year-olds have already been shown consider German *auch* ('too') in sentence interpretation (e.g., Berger & Höhle, 2012).

The current study builds up on this argumentation and aims to refine the onset age of the understanding of German *auch* ('too') and *nochmal* ('again') by considering their presupposition-triggering status in the design of the experimental task. Specifically, a simple looking-while-listening experiment was created, requiring no explicit reaction from participants. In each trial-film, children are visually and verbally introduced to an action (e.g., jumping), performed by one of two similar toy characters that are present in the scene (*Look! Jumping! Jumping is fun!*). After this, the other one applauds. Children then hear one set of two subsequent test utterances - either containing *auch* or *nochmal* (*Bingo should jump now, {too / again}. Come on, Bingo, jump, {too / again}, okay*?). To anticipate the announced action, children have to assign the correct referent to the animal name *Bingo*. Crucially, in order to do so, children must make an inference based on the presuppositions triggered either by *auch* or by *nochmal*. We expect a higher proportion of anticipatory looks to the clapping character after *auch*, compared to *nochmal*, in the timespan of 2000 ms following the offset of the last test sentence.

Results of 24 children aged 2;5-2;7, and 24 adults, each tested in two *auch*- and two *nochmal*-trials are in line with our predictions: when the test sentences contained *auch*, both children and adults expect the clapping character to do the announced action,– with the proportion of looks to this character being above chance (children: 64% / adults: 91%). When the test sentences contained *nochmal*, proportions of looks to this character are below chance (children: 35% / adults: 8%); (children: $p < .01$ / adults: $p < .001$ for all comparisons). We look also at the time course of looks during the presentation of the test sentences: Results indicate that children (as well as adults) compute the presuppositions shortly after they encounter the triggers. Our data indicate that children succeed in understanding and differentiating between sentences containing different presupposition-triggering additive expressions at least from age 2.5 years on.

Berger & Höhle (2012) Restrictions on addition: Children's interpretation of the focus particles *auch* 'also' and *nur* 'only' in German. *Journal of child language*, *39*(2). **Bergsma (2006)** (Un) Stressed ook in Child Dutch. In *Semantics in acquisition*. Springer Netherlands. **Nederstigt (2003)** *Auch and noch in child and adult German* (Vol. 23). Walter de Gruyter. **Schwarz (2007)** Processing presupposed content. *Journal of Semantics*, *24*(4). **Kim (2008)** Processing presupposition: Verifying sentences with 'only'. *University of Pennsylvania Working Papers in Linguistics, 14*(1), 17. **Cummins et al. (2013)** Backgrounding and accommodation of presuppositions: An experimental approach. *Proceedings of Sinn and Bedeutung 17*. Chemla E, Homer V, Winterstein G (Eds).

Before and after, and processing presuppositions in discourse

Ming Xiang, Emily Hanink & Genna Vegh (University of Chicago)
mxiang@uchicago.edu

The connectives “before” and “after” order eventualities in the temporal domain. Münte, Schiltz and Kutas (1998, *Nature*) showed that, compared to *after*-initial clauses, *before*-initial clauses elicited a larger prolonged left frontal negativity, as early as 300ms after the onset of the connective (e.g. *Before/After the scientist submitted the paper, the journal…*). Comprehension cost on “before” has standardly been taken to show the influence from our conceptual knowledge about how events are temporally sequenced in the real world: more computations take place for *before*-initial clauses because they present events out of chronological order, whereas *after*-initial clauses do not. In the current study, we show that the original results arise not because of this clash in temporal sequence, instead, it is an example par excellence of our rapid sensitivity to presuppositions in discourse.

Temporal *before* and *after*-clauses generally presuppose the veridicality/factuality of the eventuality they denote. For instance, the default interpretation of (1) is that John indeed won an Oscar. However, although *after*-clauses are exclusively veridical, *before*-clauses allow non-veridical events, and receive an interpretation that the denoted events are likely to have had happened, as shown in (2), where changing “before” to “after” leads to contradiction (Heinämäki,1974; Condoravdi, 2010). Comprehenders therefore couldn’t be certain whether the veridicality presupposition for *before*-clauses is satisfied until they read the main clause. The process of maintaining such uncertainty may be the source of the prolonged negativity observed for “before”, which is associated with extra working memory (WM) burden.

(1) *Before/After John won the Oscar, he bought a big house.*

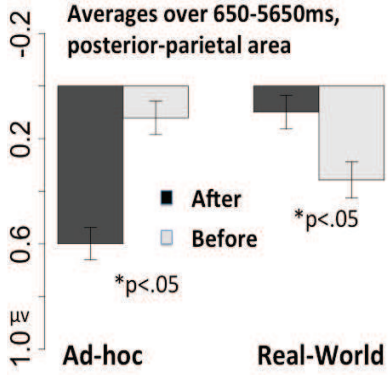
(2) *Before/*After the bomb exploded, it was defused by the police*.

We conducted an ERP experiment to test this hypothesis, using two sets of stimuli. The first set contains “Ad-hoc events”, which are arbitrary events, as exemplified by (1). In the “Real-World events” set, the temporal clauses always contain a true, widely-known, historical/cultural event, as in (3):

(3) *Before/After Star Wars came out, George became interested in astronomy.*

Since the “Real World” events eliminate any uncertainty associated with the veridicality of the *before*-clauses, we predict a reduced prolonged negativity on the Real-World *before*-clauses compared to the “Ad-hoc” ones, eliminating differences between *before*- and *after*-clauses. A hypothesis based on conceptual temporal sequence predicts no change between the two types of events.

ERPs were obtained over a prolonged 6200ms time window (with an additional 200ms pre-stimulus baseline) from the onset of the connective *before/after*, with the first 3150ms being the temporal clause, and the next 3050ms the main clause. Preliminary results from 23 native English speakers (80 Ad-hoc and 80 Real-World *before/after* sentences, and an additional 80 sentences with *when*-clauses) showed that over the 650-5650ms time window there is a ROI (anterior to parietal distribution) x Event (*Ad-hoc* or *RealWorld*) interaction in the midline area ($F(3,66)=3.98$, $p<.05$), with an Event x Connective (*Before* or *After*) interaction ($ps<.05$) in both midline-posterior and midline-parietal areas. There is also a ROI x Event x Connective interaction in the peripheral area ($F(2,44)=3.9$, $p<.05$), reflecting an Event x Connective interaction ($p<.05$) in the peripheral posterior region. Further analyses showed an Event x Connective interaction in the combined posterior-parietal region (Figure 1, based on data from 10 electrodes, $F(1,22)=5.3$, $p<.05$), with no effect from Clause (temporal or main clause). In the Ad-hoc events, there is a larger negativity for “before”, replicating Münte et al. (although with a posterior, not frontal, negativity). But crucially the effect is reversed in the Real-World events: the negativity on *before*-sentences is reduced compared to the Ad-hoc before-sentences, lending support to the presupposition-based hypothesis; at the same time, we also observe an increased negativity for *after*-sentences, possibly due to the fact that after-clauses trigger more causal implications for real-world events, hence more WM cost.



Cognitive resources underlying scalar implicature: A subclinical study

E. Matthew Husband & Nicholas Angelides (University of Oxford)
matthew.husband@ling-phil.ox.ac.uk

While the processing of scalar implicature has garnered particular interest in recent years, little research has focused on the cognitive resources required by the scalar representations such implicatures are computed over. Although implicatures may require social cognition, retrieving and processing the scales underlying the representation of scalar adjectives, for instance, may draw upon semantic memory and executive function to generate proper inferences given that their scalar representations are likely stored in semantic memory and may be accessed through controlled retrieval. To investigate these possibilities, we employed a subclinical paradigm that correlates the behavior of healthy participants with clinical measures of semantic memory, executive function, and social cognition. In sentences like (1), where *hot* entails *warm* and implicates *not scalding*, we used *cancellablity* of implicatures by an *if-not*-clause to differentiate between implicatures and entailments. Cancelling the implicature in (1a) is acceptable, whereas cancelling the entailment in (1b) and cancelling to an antonym as in (1c) is judged to be odd.

1. a. The lawyer's coffee was hot, if not scalding and too bitter to drink.
   b. The lawyer's coffee was hot, if not warm and too bitter to drink.
   c. The lawyer's coffee was hot, if not cold and too bitter to drink.

45 participants completed a two-section survey. Section 1: 26 items consisting of three scalar alternatives and one antonym adjective (9 sentences/item, pairing each of the adjectives with one another, 130 fillers). Section 2: 5 clinical questionnaires (semantic memory: Camel & Cactus Test, Concrete & Abstract Word Synonym Test; executive function: Dysexecutive Questionnaire, Frontal Systems Behavior Scale; social cognition: Autism Quotient). Participants rated the acceptability of sentences on a 1 (extremely unnatural) to 5 (extremely natural) scale in Section 1. In Section 2, participants answered the clinical questionnaires.

Acceptability ratings were analyzed using mixed effects models. Comparisons were Helmert coded into two contrasts: **Scale** (cancellation to a scale mate (1a) and (1b) vs. the antonym scale (1c)), and **Cancellation** (cancellation of an implicature (1a) vs. an entailment (1b)). Each clinical scale was added in as a fixed predictor in a separate model.

Semantic memory measures interacted positively with **Scale** (CAWT: $t = 2.698$) and **Cancellation** (CAWST: $t = 2.017$). The lower a participant's semantic memory measure, the more acceptable the cancellation of an implicature and the more unacceptable the cancellation of an entailment and the cancellation to an antonym scale. Executive function interacted negatively with **Scale** (DEX: $t = -4.444$; FrSBe: $t = -3.924$). The lower a participant's executive function measure, the more acceptable the cancellation to an antonym scale. Social cognition interacted positively with **Cancellation** (AQ: $t = 2.050$). The more autistic traits a participant reported, the more unacceptable the cancellation of an entailment.

Taken together, these data argue that while social cognition plays a role in scalar implicature, implicatures derived from scalar adjectives also draw upon semantic memory, suggesting that semantically rich adjectival scales require high functioning semantic memory to properly act as a basis for implicature calculation. Interestingly, executive function did not play a significant role in these types of implicatures, suggesting that attention/working memory are not burdened when the scales required are stored in semantic memory. Future work will extend this paradigm to quantifier <some, many, all> and ad hoc <like Greek food, vacation in Greece, speak Greek> scales which may require different cognitive resources for successful implicature calculation.

Table 1: *t*-values for the interaction of clinical score with model contrasts. *= $p<.05$, ***= $p<.001$

	Semantic Memory		*Executive Function*		*Social Skill*
	CCT	CAWST	DEX	FrSBe	AQ
Scale	0.049	2.698*	-4.444***	-3.924***	-1.495
Cancellation	0.880	2.017*	-0.828	0.299	2.050*

Indirect scalar implicatures are neither scalar implicatures nor presuppositions (or both)

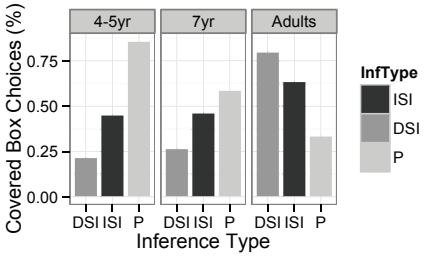
Cory Bill (Macquarie University), Jacopo Romoli (University of Ulster), Florian Schwarz (University of Pennsylvania) and Stephen Crain (Macquarie University)
cory.bill@students.mq.edu.au

Introduction: Direct scalar implicatures (DSI) have been investigated extensively in both children and adults. In contrast to adults, children have generally been found to more readily accept sentences with 'some', such as (1) in contexts that would validate stronger statements, with ''all'. A few studies have also investigated indirect scalar implicatures (ISI) such as (2), yielding a similar overall pattern of responses for children and adults (e.g., Musolino and Lidz 2006, Katsos et al. 2011). Previous studies have not, however, directly compared the two kinds of inferences. The null hypothesis, H1, is that children and adults should treat DSIs and ISIs in a uniform manner. More controversial is the status of presuppositions (P) in sentences like (3), since some (but not all) researchers have proposed that Ps are governed by the same mechanisms that are invoked in computing DSIs (and perhaps ISIs). In any case, recent findings (Bott & Chemla 2012) indicate that adults experience a processing cost in accessing an interpretation of (3) in a context in which the global presupposition is not satisfied - i.e., the bear didn't even participate in the race. This leads to H2: children are expected to be even more likely than adults to reject (3) in contexts that fail to satisfy their global presuppositions.

	Sentence	**Inference**	**Type**
(1)	Some giraffes have a scarf	-> *Not all giraffes have a scarf*	DSI
(2)	Not all giraffes have a scarf	-> *Some giraffes have a scarf*	ISI
(3)	The bear didn't win the race	-> *The bear participated in the race*	P

Experiment: 20 adults, 16 4-5 year-olds, and 14 7-year-olds were presented with sentences like those in (1-3) in a variant of the Covered Box paradigm (Huang et al. 2013). Each trial consisted of a context picture and two critical pictures, one visible and one covered. The experimenter produced a short description of the context picture followed by a test sentence that purportedly described only one of the two (visible or covered) pictures. The participants' task was to decide whether the test sentence described the visible or the covered picture. The visible picture was only consistent with the literal meaning, absent the associated inference (e.g., for (1), all the giraffes had scarves in the visible picture). Therefore, selection of the covered picture was taken as evidence that the participant had generated the relevant inference. Following a training session, there were 4 test trials and 4 controls for each condition (DSI, ISI, P).

Results & Discussion Selection of the covered picture (indicating presence of inference) depended both on age and type of inference. Children (4-5yrs & 7yrs) were less likely to draw DSIs than ISIs, and most likely to draw Ps; adults exhibited exactly the reverse pattern. Logistic regression mixed effect models revealed significant interactions between Inference Type and Age Group. Within each group, all inference types differed significantly from one another. Within Inference Type, children significantly differed from adults for all types, and 7-year olds differed from 4-5 year olds in computation of presupposition interpretations. The presupposition results for adults are consistent with the processing results of Bott and Chemla's (2012) and with traditional theories. However, the differences between the two SI conditions, for both children and adults, were not anticipated by any current theory we are aware of. And since the P responses also differ from both SIs, a three-way distinction appears to be required. We offer two possible interpretations of these results.



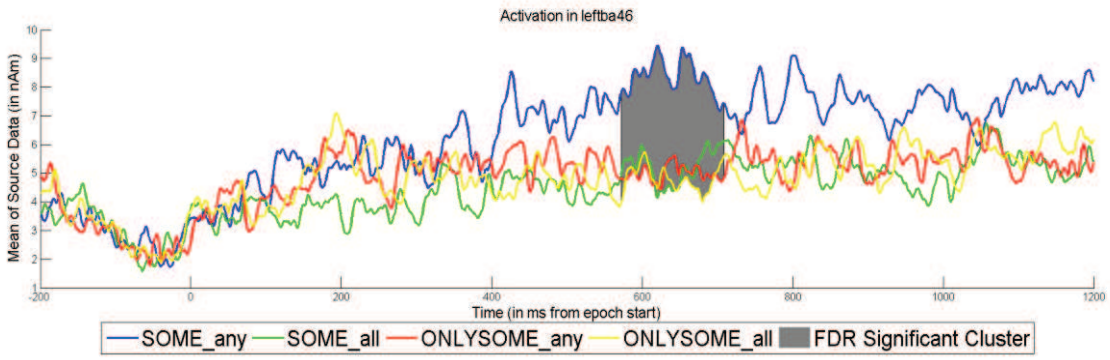
Neural correlates of realizing scalar inferences: an MEG study

Stephen Politzer-Ahles (New York University Abu Dhabi)
spa268@nyu.edu

Natural language comprehension requires inferring messages that are not part of the literal meaning of the expressions being comprehended. For example, “Some of my relatives are coming” may be interpreted as “not all of my relatives are coming”, even though the latter interpretation is not part of the literal meaning of the expression (according to some accounts). Little is known about how these inferences are cognitively realized (e.g., how easy or effortful the inferencing process is, and how inference realization relates to and interacts with semantic processes), and even less is known about what neural mechanisms subserve this process.

We used magnetoencephalography (MEG) to investigate brain activation related to realizing these types of inferences in 13 English-speaking volunteers. The experiment compared neural activation to the phrase “some of” in auditorily-presented discourse contexts in which the “some +> not all” inference is realized, versus contexts in which it is not. Inference-supporting contexts were multi-sentence vignettes in which the question under discussion is whether “all” is true (e.g.: “Mary asked John whether all of his relatives were coming; John said that some of them were”), and thus the “not all” interpretation is relevant to the discourse and likely to be derived. Inference-nonsupporting contexts were maximally similar contexts in which the question under discussion is whether “any” was true (e.g., “Mary asked John whether any of his relatives were coming; John said that some of them were”), and thus the “not all” interpretation is irrelevant and less likely to be derived. Behavioral studies using this paradigm (Breheny, Katsos, & Williams, 2006; Bergen & Grodner, 2012; Politzer-Ahles & Fiorentino, 2013) have reliably found effects of context downstream of the quantifier, but effects of context have not been reliably observed at the quantifier itself. In the present study, MEG was used to provide a more sensitive measure while participants listened to the vignettes for comprehension, and to probe the neural substrates and timecourse of the process of making inferences. Source activation was assessed over multiple anatomically-defined frontal and temporal regions of interest over a 1-second interval using temporal clustering and false discovery rate correction.

About 600 ms after the onset of “some”, middle frontal cortex (BAs 46 and 44) showed greater activation in the inference-nonsupporting than the inference-supporting condition. This effect was not observed in a pair of control conditions where the critical word was “only some” (where no inference is required to realize the “not all” interpretation). This pattern may reflect the process of making inferences when lack of contextual support makes inferencing more difficult, or it may reflect inhibiting inferences. The present results constitute some of the first evidence for neural activity that may be specifically related to realizing linguistic inferences.



Poster #2053

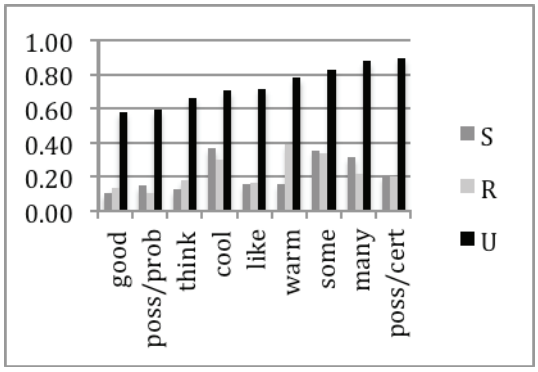
Processing Scalars

Mandy Simons (Carnegie Mellon University) & Tessa Warren (University of Pittsburgh)
simons@andrew.cmu.edu

Some argue that scalar implicatures arise from purely pragmatic principles (e.g. Geurts, 2000) whereas others argue they are a result of the compositional semantic process (Fox et al. 2013). Neither view predicts differences in the likelihood of strengthening across different scalar items. The current experiment tests for such differences, and also begins to probe how manipulations at the middle versus end of a scale affect the likelihood of strengthening.

Participants (N=43) read 36 passages and after each passage rated 7 sentences (3 critical, 4 filler) as consistent (coded as 0) or inconsistent (coded as 1) with the passage. Each paragraph contained three weak scalars (see graph below for the set), each of which was tested by a sentence whose consistency with the paragraph required either an unstrengthened reading of the scalar (conditions U and R) or a strengthened reading (condition S). Unstrengthened readings were probed in two ways: by invoking only the top of the scale (U), or a range including the entire scale (R). Critically, unlike in e.g., Grodner & Russell (2013) and van Tiel et al. (2013), these sentences did not include the scalar term or its scale-mates; instead they implemented an underlying semantic scale (e.g. in a passage about Sally feeling cool after going swimming, the sentences were: "After swimming, Sally would: (U) be blue-lipped and shivering./ (S) not be hot, but would not be blue-lipped and shivering./ (R) be anywhere from comfortably not-hot to blue-lipped and shivering.") This method was designed to avoid biasing subjects towards strengthened readings (cf. Geurts & Pouscoulous 2009). The method provides access to participants' interpretations without requiring them to explicitly articulate them.

Analyses used LMERs with crossed random effects for participants and items and extensive random structure. There was a linear effect of scalar (ordered as in graph; β=.16, z=3.07, p <.003; ANOVAs also showed a main effect of scalar: F1(8,336)=20.48, p<.001, F2(8,99)=4.75, p<.001). Across the 9 scalars, sentences invoking the scale's top (U condition) were more often judged inconsistent than the other two conditions (β=2.8, z=15.00, p <.001), but judgments for S and R did not differ (p=.68). This is interesting, given that R conditions included U conditions; participants seemed willing to accept scenarios *including,* but not *limited* to, the top of the scale as compatible with the weak scalar. There was also a marginal interaction between scalar and the difference between U and S/R (β=.12, z=1.79, p =.07). Seven of the scalars had been tested in van Tiel et al. (2013); their relative ordering by strengthening rate in that study replicated and reliably predicted rates of rejecting U in the current study (β=.23, z=3.03, p<.005). In a novel finding, manipulations of the top of a scale (i.e. possible/probable vs. possible/certain), implemented by variations in the critical sentences, had strong effects on rejecting U (β=2.98, z=4.5, p <.001), but manipulations in the middle of a scale (i.e. some/all vs. many/all) did not (p=.49). This provides further evidence that participants reject the very top of the semantic scale as compatible with the weak scalar, but take non-maximal but stronger values to be compatible.



These results indicate that there is considerable heterogeneity in rates of strengthening across scalar terms and across manipulations at different points on a scale. They further suggest that rather than ruling out the top of the associated semantic scale altogether, strengthened readings of scalars simply exclude the possibility that the relevant value is known to be at the top of the scale. This warrants a re-examination of the nature of scalar inference.

Prosody affects scalar implicature generation

Marie-Catherine de Marneffe & Judith Tonhauser (The Ohio State University)
mcdm@ling.ohio-state.edu

Utterances of sentences with a scalar expression may give rise to a scalar implicature (Horn 1972): *Some apples are ripe* may implicate that not all the apples are ripe. Recently, there have been empirical investigations on the factors that affect whether a scalar implicature arises (i.a., Zondervan 2010, Degen 2013, Goodman & Stuhlmüller 2013). However, such investigations remain agnostic about the prosody of the utterances. In (1), for example, they seem to assume that Julie's utterance is realized with a "neutral" contour (a high pitch accent on the stressed syllable of the adjective, a low phrase tone and a low boundary tone (H* L-L% in the ToBI framework, Beckman et al. 2005). In contrast to the "neutral" contour, the "uncertainty" contour (also "rise-fall-rise", Ward & Hirschberg 1985, Constant 2013) contributes to the meaning of Julie's utterance. This contour is characterized by a rising pitch accent on the stressed syllable of the adjective, a low phrase tone and a high boundary tone (L*+H L-H%).

(1) Mike: Was your date hot? Julie: He was good-looking.

In this paper, we show that the prosody of utterances is another factor that affects the degree to which a hearer believes a scalar implicature to be part of the speaker's intended meaning. We conducted a perception experiment that explored the degree to which hearers draw scalar inferences from utterances with "neutral" and "uncertainty" contours. 57 speakers of American English (recruited via Amazon's Mechanical Turk) listened to 22 dialogues between Mike and Julie, who they had been told were catching up over coffee. Participants were asked to judge whether they thought Julie intended a negative or a positive answer to Mike's question on a scale from 1 ("definitely no") to 7 ("definitely yes"). The target dialogues were created from 16 adjective pairs. In one set of dialogues (SW dialogues), the stronger adjective was in the question and the weaker adjective was in the answer, as in (1); here, answers were realized with both the "neutral" and "uncertainty" contours. Two ToBI-trained speakers recorded the stimuli. To ensure that the adjective pairs used in the dialogues are interpretable on the same scale, with one adjective stronger than the other, we first gathered acceptability judgments using linguistic tests for scalar adjectives (*My date was good-looking, even hot* vs **My date was hot, even good-looking*). In the control condition, the weaker adjective is realized in Mike's question and the stronger adjective in Julie's answer (WS dialogues, *Was your date good-looking? He was hot*); we expected these to lead to more positive answers by the participants than the SW dialogues. In the WS dialogues, Julie's answers were only realized in the "neutral" contour. There were also 6 filler dialogues in which the adjectives were antonyms (e.g., *Was your party fun? It was boring*); we expected these dialogues to invariably lead to negative answers. We ran mixed-effects ordinal regression models predicting the 7-point Likert scale, with random intercepts and random slopes for subjects and items. We found significant effects of adjective strength and of prosodic contour: SW dialogues received significantly less positive ratings than WS ones ($p < 0.001$) and, in SW dialogues, answers with the "uncertainty" contour received significantly less positive ratings than answers with the "neutral" contour ($p < 0.01$).

Our finding broadens the set of factors affecting scalar implicature generation: hearers are more likely to infer that the speaker intended a scalar implicature when using the "uncertainty" contour than the "neutral" one. This finding calls into question the suitability of using written stimuli for exploring implicature generation since such stimuli do not control for prosody. We also argue that our novel experimental paradigm, which relies on judgments of indirect answers to polar questions, is a more natural way of probing for implicatures than some previously used ones.

Pragmatic inferencing across scales: Linguistic and extra-linguistic effects

Yi Ting Huang (University of Maryland College Park)
ythuang1@umd.edu

Language comprehension is opportunistic. It makes use of multiple cues from both linguistic and extra-linguistic sources to rapidly converge on an intended meaning. For example, while the semantics of "*some of the cookies*" are compatible with the total set (ALL OF THEM), listeners often generate a pragmatic implicature to exclude competing alternatives on the lexicalized scale (NOT ALL). Critically, such inferences also occur when ad-hoc scales are created via discourse context. For example, while the meaning of "*girl with cookies*" is compatible with the superset (COOKIES AND CAKES), listeners often generate an implicature to restrict interpretation to just the subset (NOT CAKES). The flexibility of this inferential process demonstrates that scalar structures can be created via linguistic and extra-linguistic inputs. However, it raises questions about how these distinct sources of information are processed during comprehension and whether the scalar structure of one is more accessible than the other. One possibility is that frequent usage reinforces inferred interpretations for lexicalized scales, leading to rapid implicatures. Another possibility is that salient discourse context directly primes the conceptual representations associated with inferred interpretations for ad-hoc scales.

Experiment 1 (n=20) examined the time-course of implicatures for ad-hoc scales. Critical trials presented listeners with instructions like (1a) while their eye-movements were measured to displays featuring subsets ($girl_1$) and supersets ($girl_2$). Since the meaning of (1a) is semantically ambiguous, fixations to the subset will reveal when inferencing occurs ("cookies" = NOT CAKES). Control trials unambiguously referred to the superset using the unique item ("cakes"). Here, reference restriction is expected to be early. *Results* – In the critical trials, participants always selected subsets, providing evidence of implicatures. However, their on-line fixations were initially equivocal between subsets and supersets, delaying reference restriction until 500ms after the critical expression. In contrast, supersets were rapidly disambiguated after 100ms, leading to a significant effect of set type beginning from the 200ms window ($p<.05$). This pattern is consistent with previous work on lexicalized scales, demonstrating that late-emerging pragmatic inferences are preceded by initial semantic analysis.

Experiment 2 (n=36) directly compared the time-course of interpreting lexicalized and ad-hoc scales. In addition to the trials above, critical trials for the lexicalized scales presented listeners with globally-ambiguous instructions like (1b) and displays featuring novel object sets. Since the referent of (1b) is unknown, preference for subsets will reveal whether an inference was used to restrict interpretation ("some" = NOT ALL THE OBJECTS). Control trials unambiguously referred to the superset using the stronger alternative ("all"). *Results* – Across both scales, actions in the critical trials overwhelmingly favored subsets (>85%). Disambiguation was again quicker for supersets than subsets ($p<.01$). Critically, comparisons across scales revealed that when an implicature was required, fixations were delayed in the ad-hoc compared to lexicalized scales. This led to scale by set interactions beginning 300ms after critical expressions ($p<.05$). These results demonstrate that conventionalized expressions provide a powerful tool for structuring inferential processes. Nevertheless, it also suggests that across both scales, robust pragmatic inferences are initially guided by semantic analysis.

(1) a. Click on the girl that has <u>cookies</u> ($girl_1$: COOKIES, $girl_2$: COOKIES AND CAKES)
b. Click on the girl that has <u>some of the daxes</u> ($girl_1$: 2-of-4 $OBJECT_A$, $girl_2$: 3-of-3 $OBJECT_B$)

	0 MS	100 MS	200 MS	300 MS	400 MS	500 MS
Ad-hoc / Subset	-.12	-.15	-.21	-.26	-.13	-.06
Lexicalized / Subset	-.06	-.03	.02	.02	.05	.12
Ad-hoc / Superset	-.01	.01	.14	.29	.39	.43
Lexicalized / Superset	-.04	.03	.11	.21	.28	.22

Exhaustivity and polarity-mismatch

Aron Hirsch (MIT)
aronh@mit.edu

1. Overview When can an answer to a question be interpreted as exhaustive, and when can't it? This paper experimentally establishes a link between **exhaustivity and polarity**: an answer can be inferred to be exhaustive when it matches the question in polarity (i), but not when there is polarity-mismatch (ii) (Spector 2005, Uegaki 2013 i.a.; e.g. adapted from Uegaki).

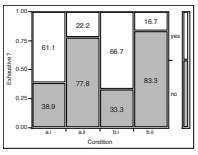
(1) There are four boys in the class: Wilbur, John, Fred, and Bill. Sue, one of the girls, is hosting a party.

(a) Which of the boys did Sue invite?
- i. She invited Wilbur.
- ii. She didn't invite Wilbur.
- iii. She **only** didn't invite Wilbur.

(b) Which of the boys didn't Sue invite?
- i. She didn't invite Wilbur.
- ii. She invited Wilbur.
- iii. She **only** invited Wilbur.

Polarity-mismatching answers can, however, be overtly exhaustified with *only*, (iii). On a theory which derives exhaustive inferences in the absence of *only* with a silent *exh* operator having a similar semantics to *only* (e.g. Fox 2007, Chierchia et al. 2009), why is *exh* impossible in (ii), while *only* is possible in (iii)? I argue that the contrast follows from processing considerations.

2. Experiment Does the strength of exhaustive inference robustly correlate with whether the polarity of the answer matches vs. mismatches the polarity of the question? Stimuli: 12 items, varying by 4 conditions, following the profile in (i) vs. (ii). Polarity of the question was crossed with polarity of the answer in a 2x2 design (run Latin Square). Two conditions had a positive question, one with a positive answer (a.i), one with a negative answer (a.ii); two conditions had a negative question, one with a negative answer (b.i), one with a positive answer (b.ii). As in (1), the dialogs were presented in a context establishing a salient set over which the *wh*-word in the question ranges. For maximum generalizability of results, the question was varied across items between object (e.g. *Which of the boys did Sue invite?*), subject (e.g. *Who came to the party?*), and VP (e.g. *What did John do yesterday?*) interrogation. The task was **binary decision**. Participants were asked to judge whether the answer to the question could be interpreted as "completely resolving" the question (yes/no), where "completely resolving" was established in the instructions to mean exhaustive. Results: Rates of yes vs. no responses by condition are plotted (24 participants), showing **significant effects** of polarity match vs. mismatch for both positive (a.i vs. a.ii, $|z|=5.68$), and negative (b.i vs. b.ii, $|z|=6.74$) questions. Conclusion: exhaustive inferences are significantly more robust when there is polarity match, whether the question is positive or negative. This can be formalized by encoding anaphoricity to a polarity-matching question as a presupposition of the *exh* operator (extending Beaver & Clark 2010).



3. *Exh* vs. *only* Beaver & Clark argue that *only*, like *exh*, presupposes a polarity-matching question. This makes the felicity of (iii) surprising. I will provide evidence that (iii) involves **accommodation** of an unasked polarity-matching question, and that similar accommodation is blocked with *exh*, accounting for the different distribution of the operators. I argue that accommodation of a question of opposite polarity to the asked question incurs a **processing cost**, which is **avoided if it can be** (economy). Cf. (ii) vs. (iii). Because ***exh*** is **silent**, the pronounced string in **(ii)** is **ambiguous** between an exhaustive parse with *exh* and a non-exhaustive parse without *exh*. The former requires accommodation of a polarity-matching question to satisfy the presupposition of *exh*; the latter doesn't. To avoid the cost associated with accommodation, the non-exhaustive parse is dominant: (ii) is preferentially interpreted as a partial answer to the asked mismatching question, rather than as an exhaustive answer to an accommodated matching question. Because ***only*** is **pronounced**, this competition never arises in **(iii)**; there is no way out of accommodation.

Focus preferences for focus-sensitive particles (and why)

Jesse A. Harris (Pomona College) & Katy Carlson (Morehead State University)
Jesse.Harris@pomona.edu

Only and *even* are both focus-sensitive particles, in that their contribution to the meaning of a sentence is modulated by the location of pitch accent in the element they modify (König, 1991). Given the amount of semantic calculation needed to integrate them into a sentence, one might hypothesize that they would be most natural in sentences with tightly controlled scope and clear focus structure, cf. Büring & Hartmann's (2001) Closeness Principle – *Focus particles adjoin as close to the focus as possible* – a principle assuming syntactic flexibility in the particle's position. Instead, three studies of the two particles' preferred positions and their interactions with focus placement find that they differ in the syntactic constraints on their placement, but both prefer prosodic flexibility as in the Broadest Focus Principle (1).

In Experiment 1, subjects rated written sentences like (2), with *even* or *only* placed immediately before the verb (*handed*), before the first DP (*sandwich*), or before the second DP (*sister*), for naturalness on a 7-point scale (7 as "totally natural"). *Even* elicited lower ratings than *only* did. For both particles, pre-verbal position was preferred and the second DP position was degraded; yet *even* was rated notably worse when after the verb. These results suggest that *even* is more syntactically constrained than *only*, and the preferred position for both particles is pre-verbal, a position which allows for more focus structures, as in (1).

In Experiment 2, subjects rated auditory sentences like (3) with *even* or *only* before the verb, and a contrastive accent on the verb (*handed*), the first DP (*sandwich*), or the second DP (*sister*) for naturalness as above. Whereas Closeness predicts a preference for verb accent, Broadest Focus predicts a benefit for accenting the last argument, consistent with narrow focus on that element or broad focus on the whole VP (Cinque, 1993). Indeed, sentences were rated highest when focal accent appeared on the second DP, and lowest when on the verb (which forces narrow focus). Also, sentences with *even* were rated higher than those with *only*, and the penalty for narrow verb focus was greater for *only* than *even*. Here again there is a difference in the level of ratings for the particles, but the same pattern within each particle's response to accent. Both particles prefer sentence-final accent, which allows for the most focus possibilities.

In Experiment 3, an auditory forced-choice completion study, subjects completed sentences like (4), with *even* before the verb and sentence-final accent, with one of two written responses. Conjuncts in *let alone* coordination must stand in focal contrast (Toosarvandani, 2010) with either NP or VP continuations, with a bias towards the latter. If *even* has little effect on focus position in these structures (Harris, 2013), and broad focus is preferred by default, then NP accent as in (4) should be consistent with either broad or narrow focus marking. Subjects chose VP continuations 41% of the time, marginally reversing the usual VP bias of such structures (Harris, 2013). Explicit accent favored the NP here, but did not disambiguate the sentences, showing that focus can project from a verbal complement (Selkirk 1984) and that this combination of particle and accent remains flexible in interpretation.

Overall, the focus particles are united by a preference for flexibility of scope and focus position, consistent with the fact that both particles relate the information in the sentence to a larger discourse which must be accommodated when the sentences are in isolation. The more constrained the scope and focus position are, the more specific and unusual the situation which must be imagined by the listener (Roberts, 1996). Thus, Broadest Focus permits the listener to avoid potentially unnecessary discourse commitments.

1. **Broadest Focus:** Prefer focus particles in positions consistent with the most focus possibilities, while respecting syntactic preferences on particle placement.
2. Anna (even/only) handed (even/only) a sandwich (even/only) to her sister.
3. Anna {even/only} handed a sandwich to her sister.
4. Anna didn't even hand her sister a SANDWICH, let alone {a bag lunch | cook her dinner}.

Pragmatic fit and the processing of Korean honorifics

Hong Mo Kang, Jean-Pierre Koenig, & Gail Mauner (University at Buffalo)
hongmoka@buffalo.edu

Most studies of how knowledge of event-types is used in reading have focused on the interaction of knowledge of event-types and object affordances (McRae & Matsuki, 2009, among many others). This paper addresses two questions: (1) whether event knowledge used in reading goes beyond event-type and object affordances and includes the relative social status of an event's participants, and (2) whether some of this event information can be inferred.

The goal of our experiment was to examine the interaction of inferred information about the relative social status of event participants and event-type information in Korean. To determine the effect of inferred social status information, we used the honorific suffix *-nim*. Referents of *–nim*-suffixed NPs in Korean are socially superior to the speaker, while referents of NPs without this suffix are not. So, in example (1), if *–nim* attaches only to the *prince*, hearers can infer that the *prince* is socially superior to the *king*, since the *prince* is socially superior to the speaker, but the *king* is not. Although hearers *can* draw this inference, the question is *whether* and *when* they draw this inference and whether they use it in processing sentences.

(1) **[** imkum-i wangca-**nim**-ul/ wangca-lul kwungkwel-eyse
King-NOM prince-**HON**-ACC/prince-ACC palace-LOC
hwunkyeha-yss-tako/wiloha-yss-tako **]** panho-ka tanenha-yss-ta
scold-PAST-COMP/console-PAST-COMP Panho-NOM assert-PAST
'Panho asserted that the king scolded/consoled the (honorable) prince in the palace.'

To answer this question, we used verbs like *hwunkyehata* 'scold' which describe actions that, typically, social superiors perform on social inferiors (hereafter, *hierarchical verbs*). Upon encountering hierarchical verbs, hearers will assume, based on their event-type knowledge, that the agent is socially superior to the patient (in (1), *king*>*prince*). Therefore, sentences that contain both hierarchical verbs and *–nim* only suffixed to objects convey conflicting information (*-nim* being suffixed only on the object suggests the agent is socially inferior (*prince*>*king*), while event-type knowledge associated with hierarchical verbs suggests it is superior (*king*>*prince*)).

We conducted a self-paced reading experiment using Korean sentences like (1) (spaces delimit presentation regions). We manipulated two variables (a) suffix-attachment (whether objects include the suffix *-nim* (*hon*) or not (*nohon*)) and (b) verb type (whether the subordinate clause verb is hierarchical ('to scold')(*hier*) or not ('to console') (*nonhier*)). We obtained two results: (a) a main effect of suffix-attachment at the modifier region of the subordinate clause ('palace' in (1)) ($F_1(1,76)=4.26$, $p=.04$, $F_2(1,18)=5.06$, $p=.04$) and (b) an interaction between suffix-attachment and verb-type at the subordinate verb region ('scold/console' in (1)) ($F_1(1,76)=4.10$, $p=.046$, $F_2(1,18)=4.14$, $p=.050$): Reading times were elevated at the modifier region when *–nim* was suffixed to the object, as predicted if readers drew the expected inference, as kings are typically superior to princes (*hon*: 586ms, *nohon*: 561ms). Reading times were also elevated *at the subordinate verb region* but only when the verb was hierarchical and *–nim* was suffixed to the object. (***hon-hier*: 619**, *hon-nonhier*: 566, *nohon-hier*: 577, *nohon-nonhier:* 570ms). The results of our experiment suggest readers quickly and effectively integrate grammatical knowledge, pragmatic inferences, and event-type knowledge. Participants drew inferences from the information conveyed by the grammatical form of the object (only the object included *-nim*) before they had a complete propositional representation for the sentence (i.e., before they read the sentence's main verb (*assert* in (1)). They then made quick use of this inference: They detected, right at the hierarchical verb region, the conflict between the inferred agent and patient's relative social status and the agent and patient's relative social status they expected, given their knowledge of the event-types denoted by hierarchical verbs.

References

McRae, K., & Matsuki, K. (2009). People Use their Knowledge of Common Events to Understand Language, and Do So as Quickly as Possible. *Language and Linguistics Compass, 3*(6), 1417-1429.

Processing asymmetries between subject-*only* and VP-*only*

Martin Hackl, Erin Olson, & Ayaka Sugawara (MIT)
hackl@mit.edu

Summary: A long-standing puzzle in language acquisition is that children up to age 6 display surprising difficulties understanding sentences with subject-*only* such as (1a) while seemingly having no difficulty understanding sentences with VP-*only,* (1b). Moreover, when they misconstrue (1a) children understand it to mean what (1b) does, (Crain et al. '94 etc.). E.g. Kermit's answer to the question *What happened?* in (1a) is judged to be true relative to the scene in (1c) and the justifications indicate that (1a) is assigned the meaning expressed by (1b).

(1) a. Only the cat is holding a flag.
b. The cat is only holding a flag.
c. 

Based on results from a series of timed inference tasks, we, first, show that adults exhibit a parallel processing asymmetry between Subj-*only* and VP-*only* (Exp.1) and, second, identify the ease with which the scalar presupposition triggered by *only* (e.g. Beaver&Clark'08) can be satisfied as a factor that governs processing difficulty of sentences with *only*, (Exp. 2-3).

Assumptions and Hypotheses: We assume that processing difficulty for sentences with *only* ('*only S*') are affected by the ease with which a suitable ALT(S) can be construed. Specifically, all else being equal, *only S* will be more difficult to process than *only S'* if construing a stronger alternative to *S* is more difficult than construing a stronger alternative to *S'*. We hypothesize, extending Fox&Katzir's ('11) algorithm for generating alternatives, that stronger alternatives to S are easier to generate if the focus associate of *only* in S is itself scalar (Exp.2) or if the context or the focus associate make plural entities salient (Exp.1 & Exp.3).

Methods: In a timed inference task participants read a set of statements (e.g. *1. The cat is holding a flag. 2. The goose is holding a flag and a balloon. 3. The frog is holding a balloon.*) on a computer screen and then click on a button labeled "*What happened?*" Upon clicking, the premises disappear and a target sentence is displayed (e.g. *Only the cat is holding a flag.* or *The cat is only holding a flag.*) whose truth participants assess wrt. the premises. Dependent measures are Accuracy and RT. All experiments fully cross Truth with Attachment Site of *only* (Subj. vs. VP), use 6 items per condition, and were conducted via the IBEX platform. All results are based on 60 pps and logit mixed models for assessing differences in Accuracy and linear mixed effects models of LogRTs for assessing differences in RTs.

Results: Exp.1 parallels Crain et al.'s ('92,'94) study. We observe no effects on Accuracy but we do observe a main effect of Attachment on RTs: VP-*only* < Subj-*only*. This parallels the acquisition data and can be explained under our approach if the premises are more likely to make plural objects salient than plural subjects. We argue that this is so because the premises mention *holding a flag and a balloon* and so introduce a plural object discourse referent. **Exp.2** is like Exp.1 except for the subject DP: target sentences are either of the form *Only one of the animals is holding a flag* or *One of the animals is only holding a flag. One* is inherently scalar and so should be a natural associate for *only* predicting that Subj-*only* should now be easier while VP-*only* should be harder. We observe a main effect of Attachment on Accuracy (Subj-*only* >> VP-*only*) and on RT (Subj-*only* << VP-*only*) and a Truth x Attachment interaction for RTs such that Subj-*only* False < Subj-*only* True while VP-*only* False > VP-*only* True. **Exp.3a,b** is parallel to Exp.1 except that target sentences used either a conjunctive subject (e.g. *the cat and the goose;* Exp. 3a) or a conjunctive object (e.g. *a flag and a balloon;* Exp. 3b) yielding target sentences of the form (*Only*) *the cat and the goose are* (*only*) *holding a flag.* and (*Only*) *the cat is* (*only*) *holding a flag and a balloon*. We predict that the placement of conjunction modulates processing difficulty st. conjunctive subjects should make Subj-*only* easier than VP-*only* while conjunctive objects should do the opposite. We observe no effects on Accuracy for either Exp.3a or Exp.3b but we observe main effects of Attachment on RT for both such that for Exp.3a Subj-*only* < VP-*only* and for Exp.3b VP-*only* < Subj-*only*.

Rapid inference of semantic information transfer rate in conversation

Mark Myslín (University of California, San Diego)
mmyslin@ucsd.edu

Recent work suggests that comprehenders rapidly adapt their expectations for linguistic input, such as probabilities of upcoming syntactic structures or boundaries of phonetic categories, converging towards the statistics of the current environment (Fine et al., 2013; Kleinschmidt & Jaeger, 2011). Less is known about comprehenders' potential *causal models* for these shifts in probabilities (Qian et al., 2012): do comprehenders attempt to infer, for example, communicative intents that gave rise to the current distribution? We investigate this question in the domain of utterance length and tempo in conversation, finding that comprehenders use these cues to rapidly infer their interlocuter's intended rate of semantic information transfer.

Previous research finds that comprehenders are able to adapt to fast speech, reaching ceiling on sentence verification tasks after 50-60 sentences spoken at twice a speaker's average rate in syllables per second (e.g., Adank, 2009). However, this line of research leaves unspecified whether comprehenders merely adapt to a fast rate of *phonetic* information transfer (with potentially relatively low semantic content), or additionally adapt to a high rate of *semantic* information transfer. In other words, it is not clear whether comprehenders infer the source of the adapted-to fast speech to be an intent to communicate ideas rapidly. If they do, they should expect an utterance of a given duration to contain more semantic information if the speaker has just produced speech with a high rate of semantic information transfer.

We investigate this in terms of responses to *wh*-questions such as *What instrument do you play?* in spontaneous conversation. 160 participants, recruited via Mechanical Turk, listened to 40 samples of telephone conversations from the Fisher corpus (Cieri et al., 2004), each containing two to three utterances followed by a *wh*-question. One group heard these samples at their normal speed, while the other group heard them sped to 120% of their original speed and with silences under 35 dB automatically truncated. Half of the participants then heard a beep cueing them to guess the original response, while the other half heard pink noise for the duration of the original response (at its original speed in all conditions; *mean* = 4.1 s, *sd* = 3.4 s), followed by the beep cueing the guess. We predicted that, if comprehenders inferred an intent to communicate ideas rapidly, rather than simply a speeded rate of syllables per second, they should expect more information to be contained in the noise-covered response following fast speech than normal speech, and therefore exhibit more variability in the guesses submitted.

On average, each item in each condition had 11.1 (*sd* = 3.2) unique guesses (e.g. *“guitar,” “drums,” “none,”* etc). We quantified the variability of these sets using Shannon entropy, which averaged 3.2 bits per set (*sd* = 0.55), and fit a linear mixed-effects model predicting guess-set entropy on the basis of (all interactions of) (i) the speed of playback, (ii) whether the noise-covered original response was played, and (iii) the duration of the original response, and included maximal random slopes by item (Barr et al, 2013). Two effects were significant. First, as predicted, (i) interacted with (ii): when participants could hear the duration of the original response (covered by noise), the entropy of the guesses was higher if the speech leading to the response was speeded. In other words, fast speech led comprehenders to infer that more information was contained in the subsequent noise-covered response. Second, simply hearing the noise-covered responses also increased the guess-set's entropy, potentially because this utterance-length information sometimes ruled out comprehenders' *a priori* most likely guesses.

These findings build on recent work by demonstrating that rapid adaptation is not limited to basic phenomena such as the identity of an upcoming syntactic structure or phoneme, but encompasses higher-order phenomena such as an interlocutor's intended rate of semantic information transfer in spontaneous conversation, which may in turn determine speech rate. This contributes to a relatively unexplored dimension of rapid adaptation research, namely comprehender inference of causal models (especially those related to *communicative intent* of the speaker) for the adapted-to linguistic distribution.

Recognizing what-is-said versus what-is-implicated

William S. Horton, Chris Schmader, & Gregory Ward (Northwestern University)
whorton@northwestern.edu

Unlike classic "neo-Gricean"[3] [4] approaches to meaning, "post-Gricean"[1] [5] accounts argue that aspects of meaning arising via Generalized Conversational Implicatures (GCIs) – *e.g.*, the use of "some" to convey "not all" – necessarily 'intrude' upon truth-conditional meaning and are incorporated as part of an enriched WHAT IS SAID. Recent experimental work,[2] though, has shown that, for some GCIs, individuals in a truth condition task could be trained to reliably distinguish WHAT IS SAID from WHAT IS IMPLICATED via GCI, suggesting that speakers retain access to an unenriched representation of WHAT IS SAID. In the present study, we use a running verbatim recognition procedure to investigate whether comprehenders' representations of WHAT IS SAID spontaneously encode GCI-enriched meanings. If so, sentences that render implicated meanings explicit should be recognized as matching the originals at rates similar to other kinds of meaning-preserving changes. But if GCIs are seen as distinct from literal meanings, comprehenders should be less likely to accept repetitions that express enriched meanings.

Seventy-nine participants recruited through Amazon Mechanical Turk saw a series of 48 experimental items containing a variety of GCIs. Each item was presented twice (interleaved with 48 filler items): First in its original form, and later in one of four possible repetitions, three of which represented changes to the original:

Verbatim repetition:	*Amelia came down with flu-like symptoms and went to see the doctor.*
Meaning-consistent:	*Amelia **developed** flu-like symptoms and went to see the doctor.*
GCI-explicit	*Amelia came down with flu-like symptoms and **so** went to see the doctor.*
Meaning-inconsistent:	*Amelia came down with **mono**-like symptoms and went to see the doctor.*

For each item, participants judged whether they had seen the same EXACT item previously in the experiment. The experimental items repeated at two intervals: Lag 1 (one item intervening between the original and repetition) or Lag 4 (four items intervening). The dependent measure was the probability of responding "Yes" to experimental item repetitions.

As shown in Table 1, participants were able distinguish unchanged from changed items, with the highest proportion of YES responses for verbatim repetitions and the lowest for items with meaning-inconsistent changes. The meaning-consistent and GCI conditions fell in between. Logistic mixed models revealed that the odds of responding YES to GCI items differed from both verbatim (b = 1.81, $p < .001$) and meaning-consistent (b = 1.38, $p < .001$) items. Moreover, the GCI vs. meaning-consistent contrast interacted significantly with lag (b = -1.19, $p < .01$).

Table 1: Proportions of "*Yes, same*" responses

	Lag 1	Lag 4	mean
Verbatim repetition	.82	.84	.83
Meaning-consistent	.79	.74	.77
GCI-explicit	.65	.69	.67
Meaning-inconsistent	.41	.49	.45

In this task, participants could reliably recognize GCI-enriched sentences as different from the originals, but their ability to distinguish GCIs relative to other types of semantically consistent changes diminished with lag. Comprehenders may initially be able distinguish GCIs from other propositional content, but distinct representations of the two may not be fully encoded into long-term memory.

[1] Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell.
[2] Doran, R., Ward, G., Larson, M., & McNabb, Y., & Baker, R. E. (2012). A novel experimental paradigm for distinguishing between what is said and what is implicated. *Language*, 88, 124-154.
[3] Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q- and R-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context* (pp. 11-42). Washington, DC: Georgetown Univ Press.
[4] Levinson, S. (2000). *Presumptive meanings: The theory of the generalized conversational implicature*. Cambridge: MIT Press.
[5] Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition, 2nd Ed*. Oxford: Blackwell.

The cognitive and neural basis of pragmatic processing: A case study of jokes

Evelina Fedorenko (MIT), Jeanne Gaulée (Wellesley College) and Zuzanna Balewski (MIT)
evelina9@mit.edu

Neuropsychological investigations of brain-damaged individuals have long suggested that the right hemisphere (RH) plays an important role in processing non-literal aspects of language (e.g., Critchley, 1962; Eisenson, 1962; Zaidel, 1985; Myers, 1999; Lehman Blake, 2005). For example, patients with RH damage experience difficulties with processing conversational inferences (e.g., Kaplan et al., 1990) – including indirect requests (e.g., Weylman et al., 1989; Stemmer, 1994) and commands (e.g., Foldi, 1987) – humor and sarcasm (e.g., Gardner, 1975), and information conveyed by prosody (e.g., Heilman et al., 1975). However, the RH contains several distinct systems that could be contributing to these deficits, including i) the RH homologues of the high-level language processing regions (e.g., Binder et al., 1997); ii) the RH subset of the system that supports social cognition (e.g., Saxe & Kanwisher, 2003); and iii) the RH subset of the domain-general fronto-parietal cognitive control, or "multiple demand", system implicated broadly in goal-directed behavior (e.g., Duncan, 2010). It is therefore difficult to determine which components of the mind that reside in the RH are the key contributors to our pragmatic abilities. This is especially true given that many pragmatic phenomena are complex and thus possibly require some combination of linguistic processing, social abilities, and generic problem-solving abilities.

We here address this question using the functional localization approach in fMRI. In each participant (n=12), we functionally identified three sets of brain regions using "localizer" tasks: i) the language regions (defined by the *sentences* > *nonwords* contrast; Fedorenko et al., 2010), ii) the regions that support social cognition / Theory of Mind (defined by the *false beliefs* > *false photographs* contrast; Saxe & Kanwisher, 2013), and iii) the domain-general multiple demand (MD) regions (defined by the *hard* > *easy spatial working memory* contrast; Fedorenko et al., 2013). We then examined the responses of these sets of regions in the RH to jokes and their literal controls matched for various lexical-level factors known to affect comprehension, as in (1), using the materials from Coulson & Williams (2005). This approach provides a powerful way to probe the relative contributions of these three systems to pragmatic processing, while a) avoiding the common problem of reverse inference in fMRI (i.e., interpreting activity in some brain region *x* as an index of some mental process, based on prior studies having found this mental process to activate brain region *x*; e.g., Poldrack, 2006, 2011), and b) yielding high sensitivity and functional resolution (e.g., Nieto-Castañon & Fedorenko, 2012).

(1) She went on a fourteen day diet, but she only lost two weeks / ounces.

The Theory of Mind (ToM) regions, including the most functionally selective RH ToM region (right temporo-parietal junction, rTPJ; Saxe & Powell, 2006), responded more strongly during jokes than during the literal control conditions (*p*s<.05). In contrast, although the language and the domain-general MD regions responded robustly to both jokes and literal controls relative to a low-level (fixation) baseline, they did not strongly differentiate between these two conditions: a few language and MD regions showed weak preferences for jokes, but if anything these preferences were stronger in the LH language / MD regions than in the RH language / MD regions. We therefore conclude that the RH deficits in pragmatic processing as well as the left visual hemifield or right ear processing advantages for various non-literal aspects of language (e.g., Coulson & Lovett, 2004; Coulson & Williams, 2005; Coulson & Wu, 2005) are more likely arising from damage to the Theory of Mind circuits rather than the RH homologues of the language regions or the RH subset of the domain-general cognitive control system.

We will discuss the broader implications of these findings for non-literal language more broadly. For example, some phenomena – like humor processing – may rely more heavily on the social cognition system while others (e.g., understanding metaphors) may primarily recruit the language or the MD system. Furthermore, these three systems may contribute differently to pragmatic deficits in neurodevelopmental disorders (e.g., Champagne-Lavau & Joanette, 2009): for example, deficits in social reasoning may contribute to problems with non-literal processing in autism spectrum disorders, but deficits in domain-general working memory / cognitive control may contribute to pragmatic deficits in schizophrenia.

The cost of pragmatic inference in the production of referring expressions

Peter Baumann (Northwestern University)
baumann@u.northwestern.edu

In language comprehension, pragmatic inference can be effortful: in reading, sentences involving implicatures take longer to process than sentences without implicatures (e.g. Hamblin & Gibbs, 2003), and in pragmatic language games, target identification is less accurate when the message involves an implicature than when it does not (e.g. Degen & Franke, 2012, D&F).

In language production, on the other hand, speakers have the possibility to reduce the effort of pragmatic inference by *over-specification*, i.e. saying more than is strictly necessary or optimal. And indeed, there is plenty of empirical evidence that speakers do make use of over-specification. Especially research on the production of referring expressions (e.g. Koolen, Gatt, Goudbeek, & Krahmer, 2011) has identified a number of properties of referents and their surroundings that can lead to a speaker making use of over-specification.

However, in research on pragmatic reasoning about referring expressions, in particular in language games (e.g. D&F; Frank & Goodman, 2012), over-specification is often ignored, and it is instead assumed that overly specific (referring) expressions incur a cost for being longer than an optimal expression (but cf. Degen, Franke, & Jäger, 2013 for supporting evidence). This optimality assumption is often built into the experimental tasks of referential language games by artificially restricting the set of possible referring expressions available to identify a referent.

In this work, we show that even in simple language games the assumption of optimal referring expressions may be too strong. We propose that in addition to a (potential) cost for longer expressions, the speaker can also incur a cost for pragmatic inference, and that a speaker will choose over-specification if the latter cost outweighs the former.

To test this proposal, we designed a language game similar to the 'speaker' conditions of (D&F, F&G): participants (*N*=96, via MTurk) saw three-panel pictures differing by the presence or absence of two accessories (e.g. hat and sunglasses) and were asked to refer to the middle panel (target). The target either had a unique feature in the context (Fig. 1, top) or shared one feature with the two competitors (Fig. 1, bottom), thus manipulating whether determining the optimal referring expression involved a pragmatic inference (Fig. 1, top, *implicature*) or not (Fig. 1, bottom, *simple*): in the *implicature* condition, the target can (optimally) be referred to by the expression *with a hat*, if it is pragmatically strengthened to mean *with only a hat*, while in the *simple* condition the same expression *with a hat* does not involve any pragmatic inference. To allow for over-specification, we replaced the forced choice of (D&F, F&G) with an open prompt. In annotating the answers, a referring expression was counted as over-specification only if both features were explicitly mentioned (e.g. *with a hat and without sunglasses*).

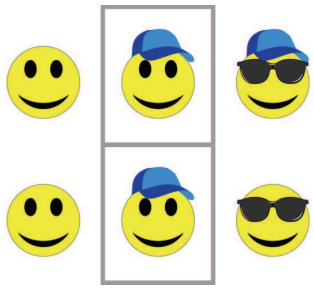


Fig. 1: Sample stimuli

Participants produced significantly more instances of over-specification in the implicature condition (46%) than in the simple condition (6%).

Our results show that even in simple language games, speakers do not always produce optimal referring expressions. Instead they make use of over-specification to avoid the effort of pragmatic inference. In terms of communicative utility, our results suggest that there is a cost for pragmatic inference, which can easily outweigh the cost for articulatory effort (or longer expressions). This proposal of a speaker cost for producing expressions involving pragmatic inference is compatible with recent evidence (Degen, Gunlogson, & Tanenhaus, 2013) showing that even for scalar terms (like *some, all*) implicatures are not as frequent as often assumed.

The effect of inferred explanations in a Bayesian theory of pronominal reference

Andrew Kehler (University of California San Diego) & Hannah Rohde (University of Edinburgh)
akehler@ucsd.edu

Background and Study Kehler & Rohde (2013) posit a Bayesian analysis of pronoun use whereby biases towards referents of pronouns ($P(\textit{referent}|\textit{pronoun})$) are determined by combining the prior probability that a referent will get mentioned next ('next-mention' biases; $P(\textit{referent})$) and the likelihood that a pronoun will be used to mention that referent ($P(\textit{pronoun}|\textit{referent})$). Crucially, the factors that condition these terms are different: Next-mention biases are determined primarily by semantically-driven factors (e.g., coherence relations), whereas the production bias is sensitive primarily to information structure and grammatical role (e.g., favoring a greater rate of pronominalizing mentions of subject referents compared to other roles; Rohde 2008, Fukumura & van Gompel 2010, Rohde & Kehler 2013).

We examine the model using data from a passage completion task. The design employs a manipulation that utilizes the fact that relative clauses (RCs) attached to direct objects can be inferred to provide explanations of the matrix event (Rohde et al. 2011). RC type is manipulated as in (1a-b):

(1) a. The boss fired the employee who was embezzling money.
b. The boss fired the employee who was hired early last year.

Although not entailed, (1a) invites the inference that the employee was fired because of the embezzling. Crucially, this inference is not necessary to make the sentence felicitous; (1b) is fine without inferring an analogous causal link between the firing and the hiring. Participants (n=17) were given a context sentence from stimulus sets including alternations like (1a-b) and asked to write a follow-on sentence (24 stimulus sets interleaved with 36 fillers). All stimuli used object-biased implicit-causality (IC2) verbs in the matrix. The continuations were annotated for coherence relation (explanation or other), next-mention (whether the matrix subject referent in the continuation was the subject of the context sentence, the object, or something else), and referential form (pronoun or other). Outcomes were modelled using mixed-effects logistic regression with maximal random effects structure when supported by the data.

Hypotheses and Results Accounts that appeal primarily to surface-level characteristics of the context (e.g. first-mention, subject assignment, grammatical role parallelism) find little to distinguish (1a-b). The Bayesian analysis does predict a difference, however, based on an interconnected sequence of referential and coherence-driven interdependencies. First, it predicts that participants will write fewer explanation continuations in (1a) than (1b), since the RC in (1a) already provides a cause (Simner & Pickering 2005; Kehler et al. 2008; Bott & Solstad 2012). This prediction was confirmed. Second, this difference is predicted to yield a difference in next-mention biases: Since IC2 verbs impute causality to the object, a greater number of explanation continuations for (1b) should lead to a greater number of next-mentions of the object as well. This was also confirmed. Third, the analysis predicts that pronoun production should be unaffected by semantically-driven factors, instead being affected only by grammatical role as discussed above. A model that crossed RC type and grammatical role showed only an effect of grammatical role. Finally, the RC manipulation is expected to affect which referent a pronoun refers to, since $P(\textit{referent}|\textit{pronoun})$ is determined in part by next-mention expectations. Analysis confirms an effect of RC type in the pronoun-only subset of the data. Since this probability is also determined in part by production biases, an effect of grammatical role favoring subjects is also predicted. This was likewise confirmed in a comparison of pronoun vs. other next-mentions.

Conclusion Biases towards referents of pronouns are sensitive to whether or not an implicit explanation can be inferred from context whereas production biases are not, revealing precisely the asymmetry predicted by the Bayesian analysis. We are currently running a second experiment with more participants to further establish the lack of effect of RC type on production biases, and with pronoun prompts in addition to full stop prompts to further establish the effects on pronominal reference.

Saturday, March 15
Poster Session III Abstracts

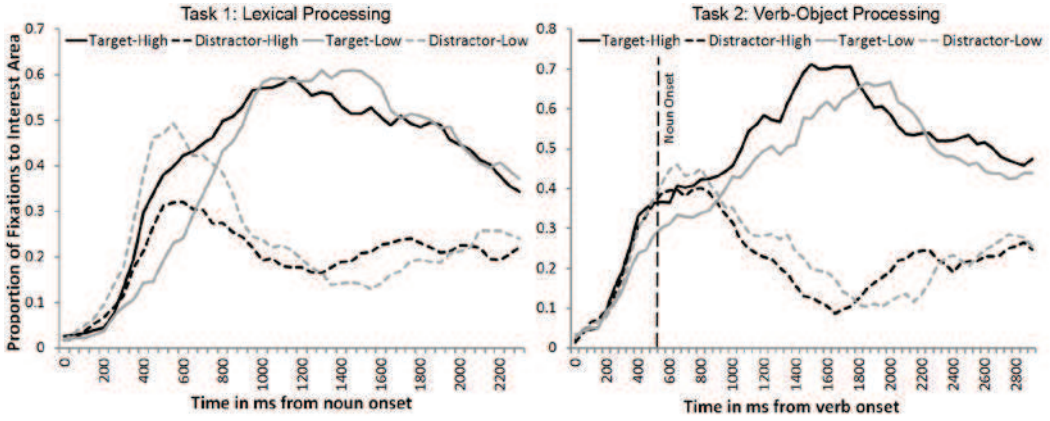
Language processing skill is not a unitary construct: Infants' vocabulary knowledge drives lexical and sentence processing

Arielle Borovsky (Florida State University), Erica Ellis (UC-San Diego), Julia Evans (UT-Dallas), Jeffrey Elman (UC-San Diego)
borovsky@psy.fsu.edu

There is increasing evidence that language-processing skills co-develop with vocabulary in infants and toddlers (e.g. Fernald, Perfors & Marchman, 2006; Mani & Huettig, 2012), though little is known about the precise nature of this relationship. One possibility is that processing skills drive vocabulary development; learners who can effectively parse the speech stream should be able to isolate and learn novel words more efficiently than those who are less fluent. Additionally, vocabulary ability may drive processing skill; individuals who have larger vocabularies may leverage this knowledge to isolate novel vocabulary in the speech stream. In the present study, we focus on this second relationship, which implies, by extension, that linguistic processing skill may vary within an individual according to the child's lexical knowledge in different category domains, while the former account predicts that processing skill should not vary according to domain knowledge.

We use a novel within-subjects approach to measure variability in real-time lexical recognition and sentence processing in individual 24 mo-old children (N=23). Rather than identifying toddlers who have "higher" or "lower" overall vocabulary skill, as in prior work, we measured each infants domain knowledge in six early-acquired categories (ANIMALS, BODY-PARTS, CLOTHES, FRUITS, DRINK, VEHICLES), via parental report on the MacArthur-Bates Communicative Development Inventory: Words and Sentences (Fenson et. al. 1994). Each infant was assigned to three "Higher" and "Lower" knowledge categories based on this report. Next, using two eye-tracking tasks, we measured infants' real-time linguistic processing of words in Higher and Lower knowledge category domains by measuring their fixations to Target images paired with an unrelated Distractor object. The first task assessed simple lexical recognition of familiar nouns (*doggie*) and the second measured target object recognition when the named object was preceded by a semantically constraining verb (*pet the doggie*!).

Our preliminary analyses focused on whether the log difference in fixation to Target and Distractor images varied according to Time (30 time bins) and Category proficiency (High vs Low) from 300-1800ms from the onset of target word (Task 1) or verb (Task 2). This analysis yielded an interaction effect of Category Proficiency and Time interacted in the familiar word recognition task, $F(30,690)=6.06$, $p<.0001$, and a main effect of Category Proficiency in the verb-object processing task, $F(30,690)=-26.4$, $p<.0001$. The results indicate that looks to target object were more robust for high proficiency items. These findings reveal that infants' lexical and sentence processing skills are not a unitary construct, as processing fluency varied according to infants' category domain knowledge. More broadly, these results suggest that language processing skills are specifically linked with the child's understanding of individual words and category domains, and not simply due to between-subject variability in general processing skill or overall linguistic experience.



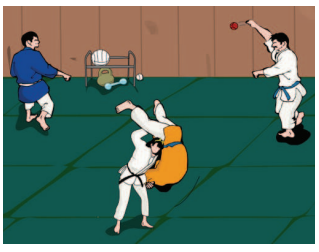
Text exposure may affect relative clause use in children and adults

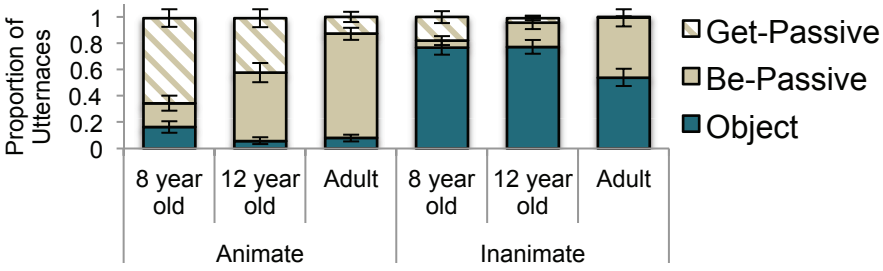
Jessica L. Montag (Indiana University) & Maryellen C. MacDonald (Univ. of Wisconsin-Madison)
jlmontag@indiana.edu

Children's difficulty with certain sentence forms, such as passives and object relative clauses, has variously been attributed to working memory limitations (de Villiers et al., 1979), limited syntactic competence (Goodluck & Tavakolian, 1982) and limited linguistic experience (Diessel & Tomasello, 2005). One way to distinguish these alternatives is to examine periods in which language input shifts dramatically, which should yield changes in child usage on an experience-based view but not necessarily on the other views. We examined one such period, when children learn to read. We focus on relative clauses (RCs) because they are hard for children, vary in speech vs. writing (Roland et al., 2007), and strongly vary in form depending on the animacy of what is described. We compared Object Relatives (ORCs: *The ball that the man threw*), Passive Relatives with a be-verb: *The baby who was carried by the man*, and get-passive Relatives: *the baby who got carried*. Study 1 investigates these structures in speech vs. text; Study 2 relates children's and adults' RC usage to their exposure to written material.

Study 1 compared distributions of the three RC types in child-directed speech (1.12 million words from the CHILDES corpus, MacWhinney, 2000) and children's literature (2.40 million words from the COCA corpus, Davies, 2008-). Analyses showed some consistencies across corpora: ORCs tend to occur with pronoun embedded subjects (*The ball I threw*), while the passive forms become more common in modifying animate entities (*The baby who was carried*…). There were also differences: the rate of RC usage was higher in writing than speech and the two genres had radically different distributions of RC types: ORCs outnumbered the two passives 96:1 in speech, but only 3:1 in written language. Though passive RCs are scarce here, previous work suggests that among all types of passives, get-passives (*the ball is getting thrown)* are more common in speech than be-passives (Carter & McCarthy, 1999). Thus reading could be a source of significant shifts in children's exposure to these three RC types.

Study 2 assessed 8-YO, 12-YO, and adults' (each N=30) reading habits via age-appropriate author and title recognition tests. All three groups participated in an RC production task, in which they answered auditory questions (e.g., "what is red?") about color illustrations (Figure 1) showing humans acting upon human and inanimate entities. Half of the trials queried animate entities, half inanimate. Pictures depicted several people and objects, inducing speakers to produce disambiguating relative clauses. Figure 2 shows that children produced a higher percentage of ORCs than adults did, an important result given claims that ORCs unduly tax children's memory or grammar. Reflecting the distributionas in both corpora, a) all three groups produced lower rates of ORCs when describing animate entities (mixed-effects logistic regression, z=11.43, p<0.001), b) ORC use decreased with age (z=-7.55, p<0.001) and c) be-passive RC use increased with age (linear trend, z=-2.69, p<0.01). Importantly, within both children and adults, individual differences in text exposure predicted production choices. High-scoring adults produced *fewer* ORCs than PRCs than low scoring adults (z=-2.39, p<0.05) while high-scoring children produced more be-passives (versus get-passives; t=3.07), on residuals of model with age. This is evidence that written language experience affects spoken production, and suggests that text exposure may contribute to differences in language use between children and adults. Implications for language production and acquisition accounts will be discussed.





Bad, or just less good? ERPs of processing Arabic agreement violations for pl. subjects.

Ali Idrissi (UAE University Al Ain) & R. Muralikrishnan (New York University Abu Dhabi)
r.muralikrishnan@nyu.edu

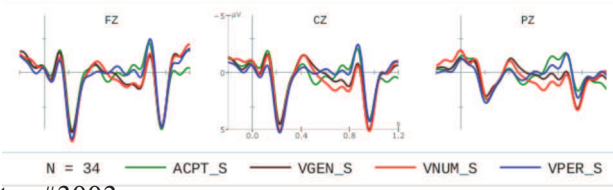
Plural subjects in Arabic are idiosyncratic when it comes to verb agreement. They require full agreement (person, number and gender) in the subject-verb (SV) order, whereas in the verb-subject (VS) order, verb agreement should be partial (person and gender only) when the subject is overt, but full if the subject is dropped. In this ERP study, we investigated whether the processing system treated violations of person- versus number- versus gender-agreement differently at the position of the verb in SV intransitive sentences. Further, we examined whether or not this pattern is qualitatively similar to that for similar violations in case of singular subjects. If qualitative differences between the two subject-types emerge for similar violations, that would indicate that the processing system is highly sensitive to the idiosyncratic behaviour of plural subjects in Arabic, even when processing sentences in the SV word-order.

Stimuli: intransitive sentences of the form Adverb-Subject-Verb-PP. Subject nouns: human (masculine & feminine) common nouns, singular or plural; Verb: either fully agreed with its subject in person, number and gender (thus acceptable) or did not agree in one of the three features (thus rendering the sentence ungrammatical; no multiple violations). Eight conditions: 2 subject types x 4 condition types. Thirty-six sentences were presented per participant in each condition; all conditions were equiprobable. Stimuli were pseudo-randomised after interspersing fillers, and presented in a rapid serial visual presentation. Participants (34 right-handed Arabic native-speakers) performed an acceptability judgement and a probe task after each experimental trial. The critical position was at the verb.

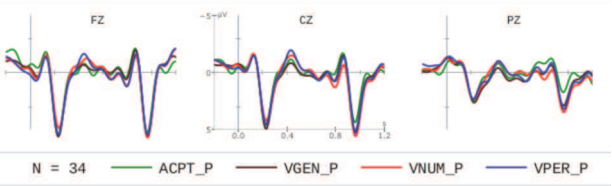
Results and Discussion: At the position of the verb, all violation conditions elicited a negativity in the 400-600 ms time-window. In the 600-800 ms time-window, there were major differences between the two subject-types. Whilst there was a general effect of condition-type in this time-window, this effect was mainly due to the interaction of subject-type and condition-type, whereby there was a graded late-positivity for the violation conditions when the subject was singular. For plural subjects, no late-positivity ensued for any of the conditions.

Given the verbs were intransitive, and the subject preceded the verb, violation of one of the agreement features constituted an anomaly, which was reflected in the negativity engendered by all violation conditions; this can be plausibly interpreted as an instance of an N400. However, major differences in processing the two subject-types began to emerge in the late-positive time-window. The graded late-positivity effect (number > gender > person) for singular subjects indicates that the number feature is the most prominent and therefore least violable. The complete absence of a late-positivity effect for agreement violations involving plural subjects, on the other hand, suggests that these were somehow syntactically less severe violations so much so that there were no differences in ERPs in this time-window between the acceptable condition and the violations. These results provide neurophysiological evidence for the fact that plural subjects are somehow special in Arabic due to their idiosyncratic agreement behaviour, and that the processing system is sensitive to factors such as these so that violations of the same sort can be utterly bad for singular subjects but not quite so for plurals.

ERPs at the verb: Singular subjects



ERPs at the verb: Plural subjects



Processing of grammatical agreement across clause boundaries

Matthew W. Lowder & Peter C. Gordon (University of North Carolina at Chapel Hill)
matt.lowder@unc.edu

The cue-based parsing framework explains key aspects of sentence processing in terms of memory processes of encoding, storage, and retrieval[1,2]. Some important questions involve understanding the types of cues that play a role in this framework and understanding how readers' ability to make use of these cues changes as memory demands increase. In the current eye-tracking experiment, we investigated the syntactic cue of grammatical number in sentences where memory search would be straightforward or more difficult.

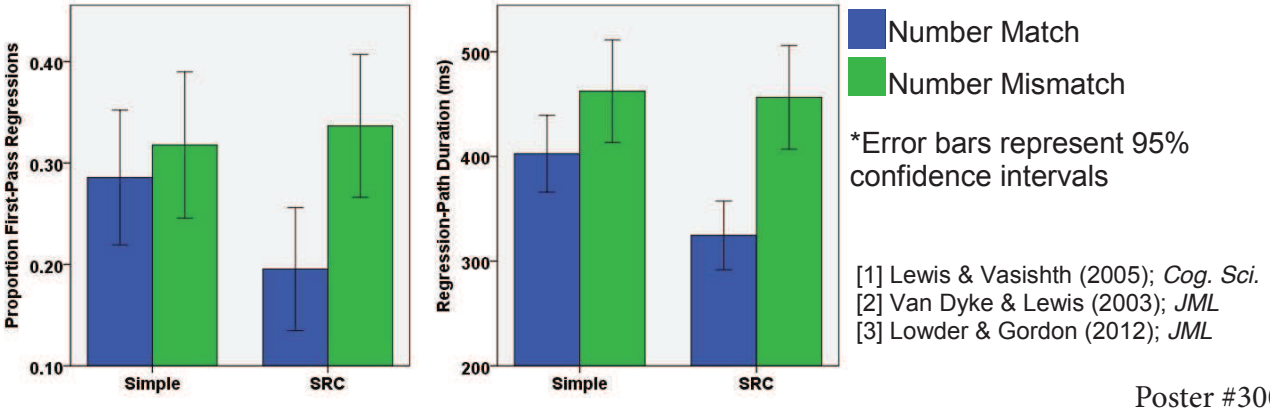
Participants read sentences that systematically manipulated number agreement between a subject NP and target verb, and whether they appeared in one clause or were separated by a subject-extracted relative clause (SRC) (see example). Participants were told that they would occasionally read sentences with grammatical errors but that they should simply try their best to understand the sentence. A comprehension question followed each sentence. Across all sentences (including fillers), 14% contained a grammatical error. Analysis of the target verb showed a higher proportion of first-pass regressions and longer regression-path durations when the subject and target verb had mismatched as compared to matched number. Critically, this effect interacted with sentence structure (regressions: $F_1 = 4.90$, $p < .05$; $F_2 = 9.46$, $p < .005$; regression-path duration: $F_1 = 5.59$, $p < .05$; $F_2 = 6.22$, $p < .02$). Both measures showed a robust effect of agreement match for the SRC condition, whereas this effect was reduced or eliminated completely in the Simple Sentence condition.

A mismatch in the grammatical feature of number caused increased difficulty in processing subject-verb relationships when the critical constituents were separated by a clause boundary compared to when they were in the same clause. In contrast, previous work shows that clausal separation of this sort reduces the processing difficulty caused by a mismatch between subject and verb in the semantic feature of animacy[3]. The current results support the cue-based parsing framework in demonstrating how syntactic cues affect processing differently, depending on the memory demands imposed by the sentence. More specifically, when readers encounter an SRC, the embedded verb triggers retrieval of an NP from the main clause to serve as the verb's missing subject. The processing difficulty caused by agreement violation in the SRC condition indicates that the grammatical number specified by the verb is an important cue in this memory search. The finding that a grammatical mismatch caused less processing difficulty in simple sentences indicates that local (within clause) subject-verb integration is less dependent on retrieval processes that rely on the grammatical feature of number.

Example sentences

The cowboy **injures** the sheriff in the bar…
The cowboy that **injures** the sheriff in the bar…
The cowboy **injure** the sheriff in the bar…
The cowboy that **injure** the sheriff in the bar…

The cowboys **injures** the sheriff in the bar…
The cowboys that **injures** the sheriff in the bar…
The cowboys **injure** the sheriff in the bar…
The cowboys that **injure** the sheriff in the bar…



*Error bars represent 95% confidence intervals

[1] Lewis & Vasishth (2005); *Cog. Sci.*
[2] Van Dyke & Lewis (2003); *JML*
[3] Lowder & Gordon (2012); *JML*

Poster #3004

Surface form effects in agreement attraction and similar phenomena

Natalia Slioussar (Utrecht University, Higher School of Economics in Moscow, and St.Petersburg University) & Natalia Cherepovskaia (St.Petersburg University)
slioussar@gmail.com

1. Background. Agreement attraction (AA) errors like (1) have recently been studied in many languages. Among other things, it was noted that only Plural local DPs cause significant AA: errors like (1) are produced more often and cause smaller delay in reading than errors like (2). It was also noted that errors appear more often in production and cause smaller delay in reading if the form of the attractor coincides with the Nom.Pl form, like in German (3a) as opposed to (3b).

(1) *The key to the cabinets are rusty.* (2) *The keys to the cabinet is rusty.*

(3) a/b. *die Stellungnahme gegen die / zu den Demonstrationen* 'the position against the$_{ACC=NOM.PL}$ / on the$_{DAT\neq NOM.PL}$ demonstrations'

Two types of accounts have been proposed for AA. (i) Subject DPs can get their number specification from a local DP due to erroneous feature percolation or similar mechanisms (e.g. Franck et al. 2002; Eberhard et al. 2005). (ii) The error arises when we access information in the syntactic structure: when we encounter a wrong verb form in comprehension or generate it, we recheck the subject DP and an attractor can interfere with this (e.g. Wagers et al. 2009).

2. Exp. 1. Nom.Pl forms of some Russian nouns coincide not only with Acc.Pl, but also with Gen.Sg forms. We compared the number of errors produced with subject DPs like (4a-c):

(4) a. *bilet na koncerty* 'ticket$_{NOM.SG}$ for concert$_{ACC.PL=NOM.PL}$'
b. *komnata dlja vecherinki* 'room$_{NOM.SG}$ for party$_{GEN.SG=NOM.PL}$'
c. *komnata dlja vecherinok* 'room$_{NOM.SG}$ for party$_{GEN.PL\neq NOM.PL}$'

In every trial, participants saw a predicate (was/were + adjective or participle), then a subject and were asked to say a complete sentence. Half of the predicates did not agree with the subjects in number. In target stimuli, subjects consisted of a head noun, a preposition selecting Acc or Gen and a local noun, like in (4a-c), but in all possible number combinations (Sg-Sg, Sg-Pl, Pl-Sg, Pl-Pl). Inanimate nouns of different genders were used.

Results. Out of all number combinations in target subject DPs, AA errors occurred only in three conditions containing DPs like (4a-c) and were distributed as follows: (4a)>(4b)>(4c). I.e. there were significantly more errors with Gen.Sg=Nom.Pl attractors than with Gen.Pl≠Nom.Pl ones.

3. Exp. 2. We found a very similar surface form effect in a different construction. In Russian, some adjective forms are ambiguous, in particular, Gen.Pl=Prep.Pl (*Prep* is *Prepositional case*). We compared case errors on nouns like (5b) and (5c) in a self-paced reading study (unlike with subject-predicate AA errors, there is no obvious way to induce such errors in production). We had 36 sets of target sentences: 12 sets with prepositions selecting Gen, Prep and Dat (control).

(5) a. *Neudachi v proshlyx sezonax zastavili komandu potrudit'sja* 'failures in previous$_{PREP.PL}$ season$_{PREP.PL}$ made (the) team work'
b. *Neudachi v proshlyx sezonov…* 'failures in previous$_{PREP.PL(=GEN.PL)}$ season$_{GEN.PL}$ …'
c. *Neudachi v proshlyx sezonam…* 'failures in previous$_{PREP.PL(\neq DAT.PL)}$ season$_{DAT.PL}$ …'

Results. In cases like (5b) (Adj$_{GEN.PL=PREP.PL}$ N$_{PREP.PL}$ or Adj$_{PREP.PL=GEN.PL}$ N$_{GEN.PL}$) the violation was detected significantly later than in cases like (5c) where morphological ambiguity plays no role.

4. Conclusions. The arguments used in the discussion of (i) and (ii) have been inconclusive so far. Our results are incompatible with (i): feature percolation is inapplicable to cases like (4b) and (5b). Approaches in (ii) need to be modified to account for our data. E.g. Wagers et al. (2009) assume that when we have a wrong verb form in subject-verb agreeement and rechecking is prompted, an attractor can provoke an error if it contains a Pl feature; the formal resemblance boosts the effect. Our data show that the attractor *need not contain* the relevant feature – it can activate the relevant feature set by virtue of being morphologically ambiguous.

Eberhard, K.M. & al. 2005. Making syntax of sense. *Psychological Review, 112,* 531-559.
Franck, J. & al. 2002. Attraction in sentence production. *LCP, 17,* 371-404.
Wagers, M.W. & al. 2009. Agreement attraction in comprehension. *JML* 61: 206-223.

Studying anaphoric dependencies using co-registration of eye movements and ERPs

Titus von der Malsburg (St John's College, Oxford), Paul Metzner (University of Potsdam), Shravan Vasishth (University of Potsdam), Frank Rösler (University of Hamburg)
malsburg@posteo.de

Eyetracking and ERPs both have unique benefits and limitations for the investigation of sentence processing. Thus, combining these methods is a natural way to amplify their potential. Recent research has demonstrated the viability of combined analyses by replicating N400 effects in response to world-knowledge violations (Kretzschmar et al., 2009; Dimigen et al., 2011). However, little is known about the potential and the challenges associated with co-registration when more subtle effects are studied. The present study focused on the resolution of anaphoric dependencies, a topic of central importance in sentence processing research. Several theories posit complexity metrics to account for dependency resolution difficulty, but it is not fully understood to what extent resolution is conditioned by dependency type. For example, it is unclear whether pronoun and ellipsis dependencies involve similar resolution mechanisms. To address this question, we conducted a co-registration experiment with four types of German sentences (English translations):

Pronoun (near / far antecedent):

The banker	...	gives Lisa$_i$ a ring		and [the lawyer]$_{masc}$ gives her$_i$ a cheap necklace.
The banker		gives Lisa$_i$ a ring	...	and [the lawyer]$_{masc}$ gives her$_i$ a cheap necklace.

Ellipsis (near / far antecedent):

The banker	...	gives$_i$ Lisa a ring		and [the lawyer]$_{masc}$ ϵ_i Anna a cheap necklace.
The banker		gives$_i$ Lisa a ring	...	and [the lawyer]$_{masc}$ ϵ_i Anna a cheap necklace.

The distance between the anaphora (pronoun or verb ellipsis) and the antecedent was manipulated using noun modifiers that followed either "the banker" or "a ring". Due to gender marking, the antecedent of the pronoun was unambiguous ("Lisa"). If both types of anaphora are processed using the same mechanisms, the distance manipulation should have a similar impact on the processing of the pronoun and the word indicating the verb ellipsis ("Anna"). Forty-six participants read 180 experimental sentences intermixed with 180 fillers. ERPs were time-locked to the onset of the first fixation on the critical word. Differences between conditions were tested using a non-parametric Monte-Carlo test (Maris & Oostenveld, 2007).

Increased distance from the antecedent elicited an early frontocentral negativity in response to verb ellipses (88ms-134ms). No ERP effects were found on pronouns, perhaps due to the high rate of skippings (60%). On the word following the pronoun, we found a negativity at frontocentral and parietal electrodes (98ms-186ms). The analysis of eyetracking measures showed that pronouns were read slower when the distance to the antecedent was long.

ERP and eyetracking data both show different effects for pronouns and ellipses suggesting different underlying resolution mechanisms. The ERP results suggest that the resolution of verb ellipses is a rapid process whereas pronouns are resolved relatively late (ERP effect on the next word). The early effect on ellipses resembles an effect found by Streb et al. (2004) in a similar experiment using rapid serial visual display and may reflect the violation of syntactic or even visual expectations (lower-case first letter instead of upper case). A methodological conclusion follows from two observations: (i) we found an eyetracking effect on pronouns but no effect in the corresponding ERPs. (ii) We found an effect in the ERPs at the ellipsis site but no corresponding eyetracking effect. This suggests that effects in one signal do not necessarily translate to effects in the other signal even when both were recorded concurrently. While this is a demonstration of the potential benefits of co-registration, the high skipping rate at the pronoun also illustrates how oculo-motor dynamics can complicate the analysis of EEG data recorded during natural reading.

Social category effects on the reconceptualization of referents in dialogue

Chris Schmader & William S. Horton (Northwestern University)
christopherschmader2008@u.northwestern.edu

Speakers routinely refer to objects in dialogue, but sometimes must select object conceptualizations that will be most informative to a listener[2] (e.g., "laptop vs. "computer"). One factor influencing this process is the listener's social category membership(s) (e.g., being a non-native speaker[1], or a casual vs. a close friend[3]). Speakers often choose referring expressions based on category-driven estimates of listeners' conceptual perspectives.

A question that arises, then, is whether the conceptualizations of objects encoded by these expressions persist post-dialogue. Previous work[4] has found evidence for lasting changes to how interlocutors conceptualize objects after negotiating reference, but these changes may not persist when speakers accommodate the (inferred) perspective of individuals from a particular social category. To address this issue, we had participants privately describe ambiguous shapes before and after referring to them with a dialogue partner. We manipulated the perceived social category of the partner (Ingroup vs. Outgroup), then analyzed whether participants would shift from their pre-dialogue descriptions of shapes (1) during and (2) after dialogue. If accommodating a less similar interlocutor requires referent reconceptualization, we expect to see stronger evidence of change in outgroup contexts, even after the dialogue.

Method. 17 pairs of undergraduates participated in one of two social category conditions: Ingroup, i.e., accurately informed they would interact with another student from their own university; and Outgroup, i.e., led to believe they would interact with a student from a local community college. Participants performed a six-trial dialogue task via Skype in which they took turns instructing each other to arrange 12 ambiguous shapes in pre-specified orders. In each trial, one participant referred to half the shapes and the other referred to the remaining half, alternating halves across trials. To obtain pre-dialogue conceptualizations of the shapes, we asked participants to independently describe each one beforehand. To obtain post-dialogue conceptualizations, we asked them to draw and label the shapes from memory after the task.

Results. For both (a) references in Trial 1 of the dialogue task and (b) post-dialogue labels, we analyzed the proportion of shapes for which participants *added* conceptual content compared to their pre-dialogue descriptions (e.g., if a participant's initial description was "teacup," then the reference "cup of tea" was coded as no additional content, whereas "teacup with handle" was coded as having additional conceptual content). Logistic mixed models with social category as a fixed effect, as well as by-participant random intercepts and by-item random intercepts and slopes, indicated that Outgroup participants' post-dialogue labels were more likely to differ from their initial descriptions (Outgroup: M = 0.62, SE = 0.03; Ingroup, M = 0.49, SE = 0.05; b = 0.57, $p = 0.05$), whereas Trial 1 references were equally likely to differ from initial descriptions across conditions (Outgroup: M = 0.82, SE = 0.05; Ingroup, M = 0.72, SE = 0.06; b = 0.72, $p = 0.17$).

Discussion. These results suggest that speakers interacting with outgroup members are more likely to change how they conceptualize objects due to negotiating reference. But rather than being driven by speakers' first attempts to refer to objects, these changes appear to result from other differences across Ingroup and Outgroup dialogues. Social context influences whether speakers alter their conceptualizations of objects due to negotiating reference in dialogue.

References

[1] Bortfeld, H., & Brennan, S. E. (1997). Use and acquisition of idiomatic expressions in referring by native and non-native speakers. *Discourse Processes*, *23*(2), 119-147.
[2] Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22,* 1-39.
[3] Fussell, S. R., & Krauss, R. M. (1989). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, *19*(6), 509-525.
[4] Markman, A. B. & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General,127(4)*. 331-354.

Speaking in context: Discourse shapes incremental preparation of simple sentences

Agnieszka E. Konopka (Max Planck Institute for Psycholinguistics)
agnieszka.konopka@mpi.nl

In written and spoken discourse, speakers normally generate sequences of pragmatically related sentences. As a result, sentence content and form can be shaped by discourse continuity and by the given-new status of the information that speakers choose to express. However, the implications of discourse context for the *timecourse* of online sentence formulation are largely unknown.

Studies examining formulation of sentences out of context (e.g., descriptions of events like *The dog is chasing the mailman*) have shown that speakers tend to look at and encode event characters in the order of mention. The timing of gaze shifts from one character to another at picture onset (0-400 ms) may vary: rapid gaze shifts to the subject character suggest radically incremental planning (Gleitman et al., 2007) while delayed gaze shifts to this character suggest that speakers may first encode information about the event as a whole (Griffin & Bock, 2000). This study tested whether a preceding context supports one type of incremental planning over the other and whether these effects depend on the ease of expressing a message linguistically.

Native Dutch speakers performed an eye-tracked picture description task (n=48 and 24 in Exp.1 and 2). Target trials showed pictures of transitive events. On prime trials, speakers read sentences describing an action performed by the agent or the patient in the upcoming target event, changing their given-new status in the discourse. The primes were pragmatically related or unrelated to the target event (confirmed by norming): related primes provided a supporting context, while unrelated primes described a plausible but unrelated event. To manipulate the ease of generating transitive target sentences, Exp.1 and 2 used intransitive and transitive (active) primes respectively. Analyses examined the effects of context on sentence form and on the timecourse of formulation (eye movements) for active target sentences.

		agent primes	*patient primes*
Exp.1:	Unrelated:	The *dog* was drooling.	The *mailman* was humming to himself.
Exp.1:	Related:	The *dog* started barking.	The *mailman* walked up to the house.
Exp.2:	Unrelated:	The *dog* licked its paws.	The *mailman* carried a city map.
Exp.2	Related:	The *dog* heard the gate open.	The *mailman* tried to drop off a package.

Sentence form. Target sentences began more often with the "given" character than the "new" character. This effect was smaller after related primes, suggesting a weaker influence of lexical accessibility on sentence form in supporting contexts than in neutral contexts.

Timecourse of formulation. Gaze durations on the subject character varied with their given-new status: speakers spent less time fixating "given" characters than "new" characters. More importantly, context also influenced the timing of gaze shifts to the subject character after picture onset. In Exp.1, speakers looked *earlier* and longer at subject characters ("given" and "new") after related than unrelated primes, suggesting that context supported radically incremental encoding. However, in Exp.2, speakers directed their gaze *later* to "given" subject characters after related primes, showing earlier encoding of information about the event as a whole in a supporting context. The difference between Exp.1 and Exp.2 suggests that the availability of a transitive sentence frame (Exp.2) can facilitate early encoding of information about the event. Thus formulation may begin with encoding of event information when both events *and* sentences are easier to encode, indicating a high degree of compatibility between formulation of a message and generation of linguistic material to express this message.

References

Gleitman, L., et al. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language, 57*, 544-569.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11,* 274–279.

Memory for what was said in conversation: Speakers and Listeners differ

Si On Yoon & Sarah Brown-Schmidt (University of Illinois, Urbana-Champaign)
syoon10@illinois.edu

According to models of language use, speaking and listening draw on the same core representations, supporting priming (Pickering & Garrod, 2004). In contrast, research in *the memory tradition* shows that speakers and listeners form distinct memories. In free-recall of unscripted conversation, memory of what was *heard* is better than what was *said* (Stafford & Daly,1984). Such findings are surprising given the well-established *generation effect* whereby words *spoken* are remembered better than words *heard* (Marsh, et al., 2001). Inconsistencies aside, this literature, however, does not speak to the question of the link between memory for language and language use. Thus, we take a different approach, aiming to link the relationship between memory for discourse referents in dialogue, with referential form.

Exp1 explored the memory contributions to the referential phenomenon *lexical differentiation* (Van der Wege, 2009) in which speakers contrast locally unique referents with similar referents from past contexts. Differentiation from past context occurs less often (~20%) than when contrasting pairs are in the same context (~90%). We ask if this is due to poor memory for past referents. Two participants (n=72) were assigned the roles of speaker (S) or listener (L). They viewed 4 pictures from different categories (e.g., never two shirts on the same display) on separate screens. The target was highlighted on S's screen. In the **contrast condition (C)**, S described a "contrast" item (e.g., checkered shirt) during 6 entrainment trials, and the target (striped shirt) at test. In the **non-contrast condition (NC)**, S described an unrelated item (e.g., towel) 6x, and the target at test. In the **contrast-but not mentioned (CNM)** condition, S described the contrast using a locative phrase (e.g., *top right one*) 6x, and target at test. L clicked each referent following S's instructions. Then, S and L did a memory test: trials showed 2 pictures from the same category (one old, one new). S&L were to click the old picture.

E1 RESULTS: Speakers differentiated locally unique referents with modified noun phrases when the referent had been discussed 21.9% (C) and 20.8% (CNM) significantly more often than when it had not 14.1% (NC). The lack of a difference between the C and CNM conditions shows that the locus of *the differentiation effect* is distinguishing referents--not their labels. The low differentiation rates were not due to poor memory: A significant interaction between role (listener vs. speaker) and type (contrast vs. target) was due to high memory for contrasts for both S and L (>95%), but significantly better target memory for speakers (L: 86.5% vs. S: 94.9%). Condition effects (C vs. CNM vs. NC) on memory were not significant for S or L.

In **Exp2** participants (n=96) alternated speaking and listening. We used a single entrainment trial and manipulated lag between entrainment and test (1 vs. 10 trials). Replicating Exp1, regardless of lag, S used more modifiers in the C and CNM conditions than the NC condition (23.4% vs.19.4% $p<.05$); these low differentiation rates were not due to poor contrast memory (>80%). Again, S remembered items better than L ($p<.05$). S memory was also more strongly influenced by naming: S remembered contrast items that were named better than items that were identified with locatives (CNM vs. C, $p<.05$). L's memory did not differ by naming.

CONCLUSIONS: (1) Speakers differentiate to distinguish current from past referents, not to avoid giving two tokens the same label (cf., Van der Wege, 2009). (2) Forgetting past referents is not the cause of infrequent differentiation. (3) Consistent with the *generation effect* (Marsh, et al., 2001), speakers remember discourse referents better than listeners. (4) While naming was not necessary for referential differentiation, it was helpful for future memory. Within *the language tradition*, understanding relevant memory processes is crucial for fleshing out a theory of how discourse history guides language use. The fact speakers and listeners *differ* in memory for what is said suggests that (a) partners are unlikely to achieve perfect coordination (cf., Pickering & Garrod, 2004); (b) a prerequisite to implicating cognitive biases or failures in the insensitivity to context, is showing that context was encoded and remembered in the first place.

Continuous acoustic information trickles up; discourse information trickles down

Sarah Brown-Schmidt & Joseph Toscano (University of Illinois)
sarahbrownschmidt@gmail.com

Recent studies have shown that differences in continuous acoustic-phonetic cues result in gradient activation of candidate lexical items during spoken word recognition. For example, 5-ms differences in voice onset time (VOT) along /b/-/p/ continua in words like *bear* and *pear* result in gradient changes in activation for the two words (McMurray, Tanenhaus, & Aslin, 2002) that are maintained across multiple phonemes (McMurray, Tanenhaus, & Aslin, 2009).

Here, we distinguish lexical from referential representations by examining pronoun interpretation for continua varying between *she's* and *he's*, and we ask if continuous acoustic changes result in graded activation of potential referents. Seven-step continua were created from *he's* and *she's* recordings by manipulating the /h/-/ʃ/ frication amplitude above 1.3 kHz.

In Exp. 1, MTurk subjects (N=50; native English speakers) listened to the continua and the unedited endpoints, and responded with whether they heard the stimulus as *she's*, *he's*, or neither. Listeners showed typical categorization functions with more *she's* responses at the /ʃ/ endpoint (z=16.92; Fig. 1), confirming the amplitude manipulation created clear /h/-/ʃ/ continua.

In Exp. 2, the pronouns were spliced into 32 stories based on Arnold et al. (2000), where different gender characters are introduced and one is referred to with a pronoun. For example:

> "*Daisy is playing the cello for Donald, as the telephone is ringing.* ***He/She*** *is looking at the phone, but it doesn't look like anyone is going to answer it.*"

Question: "Who is looking at the phone?". Answers: Donald / Daisy / Neither.

Nine versions of each story were created from the seven steps along the /h/-/ʃ/ continua and the two unedited endpoints. MTurk subjects (N=227) listened to one version of each story plus 35 filler stories. On each trial, the subject heard one of the stories and answered a question that probed interpretation of the pronoun by querying about the referent (as in the example above).

As in Exp. 1, there was a main effect of continuum step (z=22.47; Fig. 2). Moreover, responses were not strictly categorical, suggesting continuous information was maintained at the level of the referent interpretation several seconds after the stimulus. Discourse also exerted an effect, with participants more likely to encode the 1st-mentioned character as the referent (z=3.94). This persisted even at the *opposite* continuum endpoint (*z's*>2.9, *p's*<.01). That is, if *Daisy* was mentioned first, subjects were more likely to interpret the pronoun as *she's*, even at the /h/ endpoint, which was reliably categorized as *he's* outside the discourse context.

An in-progress visual-world eye-tracking study examines whether these graded representations are maintained over time, facilitating revised interpretation of the pronoun (*she→he*), and consequently the referent (*Daisy→Donald*) following disambiguating information.

Our findings allow for two important conclusions. First, they show that graded acoustic differences not only guide consideration of isolated words (e.g., *she/he*), but they also lead to graded activation of referential candidates (*Daisy/Donald*). Second, our results go beyond previous work showing effects of discourse on ambiguous stimuli (Rohde & Ettlinger, 2012) by showing that discourse can override a clear acoustic signal pointing to a different interpretation.

Fig. 1

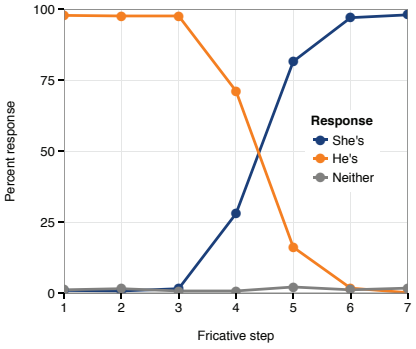
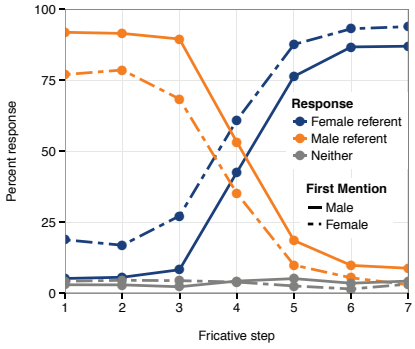


Fig. 2



The influence of partner-specific memory associations on picture naming: A failure to replicate Horton (2007)

Sarah Brown-Schmidt (U. of Illinois) and William S. Horton (Northwestern University)
sarahbrownschmidt@gmail.com

A central question regarding common ground is how this joint knowledge is stored in memory. According to the ordinary-memory view (Horton & Gerrig, 2005), associations that develop between individuals and information afford sensitivity to common ground-like representations. One finding in support of this view (Horton, 2007, Experiment 1) is the focus of the present research.

Horton (2007, E1) created situations where participants associated a label such as "banjo" with one partner, and then later named a series of pictures, including a picture of a banjo. During the naming task the partner associated with the term "banjo", or a partner not associated with that term, was seated next to the participant as they named the pictures. Speakers were 86ms faster to name pictures when the person sitting next to them was associated with the picture. Speakers were also overall faster to name pictures with familiar vs. unfamiliar labels.

The original goal of the present research was to replicate and extend the same-partner benefit in object naming. The original experiment by Horton (2007) had 16 participants who each participated in 3 within-subjects conditions during object naming: same-partner (the object label was associated with the present partner), different-partner (the object label was associated with the other partner), and new pictures. The estimated effect size for the critical partner effect (same vs. different) was d= 0.68. Based on this effect size, it should take 12 participants to reach 80% power and 42 participants to reach 99% power in our replication attempts.

E1: This was a replication of Horton (2007) E1 with materials created by the first author. Participants (n=14) first associated 16 labels (e.g., "BANJO") with partner A, and 16 different labels with partner B (e.g., "HARP"), and then in Phase 2, named pictures while seated next to A or B. **Results**: Explicit recall for partner-label pairings was high (90% for targets), but associations between partners and labels did not facilitate naming: Participants were only 3ms faster to name pictures in the same vs. different-partner condition, $t1(13)= -0.12$, $p=.91$; $t2(30)= -0.86$, $p=.40$. The effect of label familiarity was significant: target pictures were named 117ms faster than novel controls, $t1(13)=4.22$, $p<.01$; $t2(61)=1.74$, $p=.09$. **Conclusion**: Participants remembered who they studied the labels with, but partner associations did not influence naming.

E2-3: The goal of E2 was to attempt a higher powered direct replication of Horton (2007) using the identical stimuli as the original study. E3 was a conceptual replication identical to E2, except that partners A and B were salient dolls (a large red dinosaur and a Raggedy Anne doll); the goal of E3 was to test whether distinctive, but non-sentient partners would be sufficient to generate the partner-association effect in naming. **Results:** Participants (E2: n=49; E3: n=48) completed the same task as E1. In both E2-E3, explicit recall for partner-label pairings was high, and pictures with familiar (vs. unfamiliar) labels were named significantly faster ($ps<.01$). However, participants were not faster to name pictures when seated next to the partner with whom the picture label had been associated. **E2**: Participants were 26ms *slower* to name pictures with labels that had previously been studied with the same partner in Phase1, vs. a different partner ($ps>.4$). **E3**: Participants were 7ms *slower* to name pictures that had previously been studied with the same doll in Phase1 vs. a different doll ($ps>.35$).

DISCUSSION: These findings question the possible role that partner-specific memory associations have in shaping speakers' access to lexical information. In particular, E2, which was designed as a direct replication with methodological fidelity that approached the original, was run at 99% power, and provides strong evidence against the effect, if real, being as large as originally estimated. The results are reminiscent of other findings of inconsistent effects of environmental context on memory (Eich, 1985; Godden & Baddeley, 1980). Crucially, though,, the present findings should not detract from a growing body of evidence that mechanisms of memory encoding and retrieval play an important role in shaping partner-specific language use.

Representing multiple instantiations of an object:
Effects of visual and linguistic context on real-time event processing

Xin Kang, Gitte Joergensen, Gerry Altmann (University of York, UK)
xk519@york.ac.uk

As objects undergo change from one state or location to another, we must maintain and keep track of multiple instantiations of the same object (e.g. Altman & Kamide, 2009). For example, in "*The woman will drop the ice-cream. But first, she will look at the ice-cream*", we must maintain two distinct representations of the ice cream – before and after it was dropped. In three eye-tracking experiments, we used the visual world paradigm to explore how such representations interact as language unfolds.

In Experiment 1, participants (N=64) listened to sentences like 1-2(a) while viewing an image depicting the *same* ice cream in two different states (Fig A). We found participants predominantly looked more towards the intact ice-cream during the final noun phrase (*"the ice-cream"*), thus retrieving the appropriate states of the ice-cream as required by language. However, it is unclear if they treated the two ice-creams as depicting the same object in different states, or whether eye-movements reflected semantic overlap between the depicted ice-cream and its current mental representation – one that was not necessarily 'bound' to the identity of the depicted version. In Experiment 2, we displayed two *different* ice-creams (Fig B). We predicted that if the effects in Experiment 1 were just due to semantic overlap we would see the exact same pattern despite the obvious differences between the ice-creams. However, the results suggested that eye-movements in Experiment 1 were likely guided both by ascribing the same identity to the two versions of the ice-cream and by semantic overlap.

In Experiment 3, participants listened to sentences 1-(2a) or 1-(2b) and saw Fig A or Fig B. We replicated the earlier patterns of eye movements. In the Fig A condition, we predicted that if there is binding we should see fewer or delayed looks to the intact ice-cream at the end of the sentence in (b) compared to (a). This is indeed what we observed. In the Fig B condition, however, we found no difference in looks to the intact ice-cream between (b) and (a). The time-course of the effects lead us to conclude that "another" disambiguates the ambiguity in respect of whether the ice-creams in Fig A are intended to be the same or not; once this ambiguity is resolved, the two ice-creams are interpreted as distinct.

Together, these findings suggest that participants can bind together concurrent depictions of an object as reflecting the same object in different states, even though in the real world the object can only be in one state or another at any one time. We would argue that this requires the overriding effect of language – linguistic context overrides the usual requirement for temporal asynchrony between the distinct states of an object.

Example

1. *The woman will drop the ice-cream.*
2. *But first, she will look at (a) <u>the</u> (b) <u>another</u> ice-cream.*

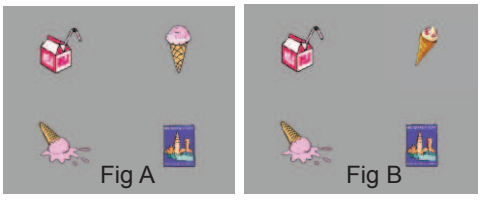


Fig A

Fig B

Reference

Altmann, G.T.M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, 111, 55-71.

Verbs drive the bus: An ERP study on the role of verb bias and plausibility in the resolution of DO/SC ambiguity in English monolinguals and Spanish-English bilinguals.

Patricia E. Román, Nicholas R. Ray, Carla Contemori (Penn State University), Edith Kaan (University of Florida), & Paola E. Dussias (Penn State University)
per3@psu.edu

Past behavioral research examining the role of verb bias and plausibility in sentences with direct object (DO)/sentential clause (SC) temporary ambiguity has shown that when verbs are strongly biased toward a DO or SC interpretation, verb bias is prioritized over plausibility [1]. Our study has two goals. The first is to use the N400 and P600 ERP components to examine whether verb bias outweighs plausibility information. The design of our study allows for this because it overcomes a limitation inherent in past work. Because past experimental designs have not completely crossed verb bias and plausibility, comparison across conditions have involved entirely different stimuli. Here, we have created stimuli that are identical across conditions (see Condition 1 through Condition 4 below), allowing for direct comparisons across conditions. A second goal is to investigate whether bilingual speakers prioritize the same information as monolinguals when reading in their L2.

1-DO verb/plausible DO (The driver heard **the passengers** could start to get annoyed)
2-DO verb/implausible DO (The driver emphasized **the passengers** could start to get annoyed).
3-SC verb/plausible DO (The bus driver worried **the passengers** could start to get annoyed).
4-SC verb/implausible DO (The bus driver claimed **the passengers** could start to get annoyed).

Twenty-four English monolinguals and 22 Spanish-English bilinguals read 124 experimental sentences and 124 filler sentences presented one word at a time on a computer screen. Verbs were strongly biased to be followed by a noun phrase (NP) functioning as a direct object or by a sentence clause. Strictly speaking, if verb bias guides the initial interpretation of the syntactically ambiguous NP (the passengers), monolingual speakers are not expected to show N400 effects at the NP when it is implausible (C2 and C4). In addition, sentence with DO bias verbs but not sentences with SC bias verbs should show a P600 effect at the disambiguating region (could start). Bilingual speakers, on the other hand, might be expected to prioritize plausibility over verb bias if they have not learned the verb biases in their L2.

Our findings indicate that the monolingual group showed no plausibility effects in the N400 time window ($F < 1$). The P600 associated with re-analysis appeared at the disambiguation region in DO bias sentences compared to SC bias sentences. Furthermore, a P600 effect arose for DO-plausible sentences compared to DO-implausible ones, suggesting that participants more easily relinquished an initial analysis when the NP was implausible (all *ps* < .05). Like monolinguals, bilinguals did not show a plausibility effect ($p > .05$). A P600 effect was significant when DO sentences were compared to SC sentences. Interestingly, bilinguals showed an N400 at the NP for SC-plausible sentences compared to DO-plausible ($p < .05$).

In all, we replicate the results from behavioral studies on the predominant role of verb bias in guiding the initial interpretation of the ambiguity. Our findings also suggest that plausibility information comes into play, but only later in processing (DO-implausible at the auxiliary verb). The results also suggest that bilinguals are sensitive to verb bias in their L2, as indexed by the presence of a P600 in DO sentences. However, the presence of the N400 in SC sentences likely indicates that they may be influenced by properties of the L1 (i.e., Spanish obligatorily requires the presence of the complementizer 'that' preceding SC clauses). Our findings argue for a constraint satisfaction account of ambiguity resolution [2].

[1] Garnsey et al. (1997)*, JML, 37,* 58-93. [2] MacDonald, & Seidenberg (2006). In M.J. Traxler & M.A. Gernsbacher (Eds.), Handbook of Psycholinguistics, 2nd Edition, pp. 581-611. London: Elsevier, Inc.

Presuppositions in the Scope of Quantifying Expressions: Eye Tracking Data

Sonja Tiemann (University of Tübingen, SFB 833)
sonja.tiemann@uni-tuebingen.de

Introduction. Presuppositions are a component of meaning displaying distinct behavior from asserted content. A longstanding debate in the theoretical literature on presuppositions concerns the question of how sentences with a presupposition are interpreted when one of the presupposition triggers' arguments is bound by a quantifying expression. The theories proposed in Heim (1983) and Beaver (1992) are on the two extreme ends with regard to what a sentence like (1) might presuppose: (1) Every nation cherishes its king. Whilst Heim's theory predicts a **universal presupposition** (that every nation has a king), an **existential presupposition** is expected according to Beaver's account (that at least one nation has a king). Crucially, both theories do not discriminate between different quantifiers or presupposition triggers. Using eye tracking in reading, we investigated the processing of German sentences with the presupposition triggers *wieder* ('again') and the definite article in the scope of the quantifying expressions *jeder* ('each') and *eine/r* ('one').

Methods. We treated every quantifier-trigger combination (*jeder-wieder*, *jeder*-definite; *einer-wieder*, *einer*-definite*)* as one separate experiment with the two-level factor **Context** ('two of three' / 'three of three'). The context varied in how many individuals possessed the presupposed property: Two out of three **C1** or three out of three **C2**.

C1: Sabine, Inge und Karin nehmen an einer Konferenz teil. Sabine und Inge haben neulich einen Laptop von ihrem Arbeitgeber bekommen, während Karin erst einen Laptop kaufen muss.
Sabine, Inge and Karin are at a conference. Sabine and Inge got a laptop from their employer recently, whereas Karin still has to buy a laptop herself.
C2: Sabine, Inge und Karin nehmen an einer Konferenz teil. Sabine und Inge haben neulich einen Laptop von ihrem Arbeitgeber bekommen, während Karin selbst einen Laptop kaufen musste.
Sabine, Inge and Karin are at a conference. Sabine and Inge got a laptop from their employer recently, whereas Karin had to buy a laptop herself.

Target: Heute hat {jede/eine} der drei Konferenzteilnehmerinnen ihren Laptop in einer Sitzung benutzt.
Today has {each/one} of.the three conference.attendees her laptop in a session used

In total, 28 items per experiment were designed. These were intermixed with 56 filler items (half of them being the items with the second quantifier and presupposition trigger). Sentences with a non-presupposing expression in place of the presupposition trigger were inserted as controls (indefinite article instead of the definite article and adverbs instead of 'again'). Based on similar eye tracking experiments (Schwarz and Tiemann 2013), increases in reading times are expected for sentences in contexts that are inconsistent with the presupposition as soon as the content of the presupposition is evident (i.e. on the critical word). In our material, this was always the word after the trigger (*Laptop* in the example above).

Results. For sentences with both triggers – *again* and the definite determiner – mean first pass durations on the word after the critical word (*in* in the example above) were significantly longer in the *two of three* context than in the *three of three* context when the quantifier was *each*. No significant differences for all eye tracking measures were present when the quantifier was *one*, and in the control conditions. Mean first pass duration times on the word after the critical word are presented in table 1.

Discussion. The results show that the interpretation of presuppositions in the scope of a quantifying expression hinges on the quantifier chosen. While sentences with *each* result in a universal presupposition, this does not seem to be the case with *one*. In the two theories discussed, the respective interpretations are independent of the quantifier. However, any theory that treats the interpretation of presuppositions in the scope of quantified expressions as a unified phenomenon, cannot account for the reading pattern obtained in this experiment. The results thus suggest that the projection behavior of presuppositions is highly dependent on the respective quantifier.

Trigger	*again*				definite det.			
Quant.	*Each*		*One*		*Each*		*One*	
Context	*two/three*	*three/three*	*two/three*	*three/three*	*two/three*	*three/three*	*two/three*	*three/three*
First Pass (ms)	228.8	204.9	231	219.7	233	209.8	218.7	222.6

Table 1. Sig. differences: *each* – *again* ($|t| = 2.48$, $p < .05$) ; *each* – definite determiner ($|t| = 1.988$, $p < .05$)

Animacy and the active construction of filler-gap dependencies in relative clauses

Emily Pendleton & Matt Wagers (UC Santa Cruz)
ependlet@ucsc.edu

Object relative clauses (ORC) are more difficult to process than subject relative clauses (SRC), but inanimate relative clause heads modulate or abolish this difficulty [1-4]. The observations behind this generalization typically reflect reading time on the verb [1,4], sometimes but not always after both potential NP arguments have been read [2, cf.3]. However, the source of ORC difficulty in general stems independently from the verb as well as from the relative-clause internal subject NP [5]. The latter source of difficulty is related to the fact that comprehenders preferentially expect empty subject positions inside relative clauses. Such an expectation may directly reflect statistical properties of the language [6] (SRCs being more prevalent than ORCs); or the mechanism of a parser that prefers to license grammatical features early [7]. Here we test directly whether animate RC heads alter the expectations that are (dis)confirmed at the RC subject itself. We find that only animates predict empty subject positions (i.e., gaps).

Design. To test the effect of animacy on predictions at the subject phrase, we use the filled-gap paradigm [8]. We compared relative clauses in which explicit predictions for a subject gap *can* be made to ones in which an explicit prediction *must* be made for a filled subject position. Reading times can be measured at an overt subject NP which foils the first prediction and confirms the latter. While [8] found no reliable subject filled gap effect, [9] showed that, when an RC initial adjunct was inserted, a subject filled gap effect emerged. Following [9], we constructed 32 sentences in a 2 × 2 design, crossing FILLER CATEGORY [NP v. PP] with RC head ANIMACY. The filler category manipulation allows us to contrast a possible gap prediction (FILLER:NP) with a definite prediction for an overt subject (FILLER:PP). In the example below, the critical subject NP is highlighted in bold and the filler category contrast is in curly braces. Sentences were read word-by-word in self-paced reading, followed by comprehension questions. 34 persons participated for course credit.

- ANIM	The school principal brought in the textbooks	{ which | for which } …
+ANIM	The school principal brought in the new student	{ who | for whom } …
	...according to the annoyed teacher, **the class** had been waiting (for) all morning.	

Results. At the critical RC subject NP, there was a significant interaction between ANIMACY and FILLER ($t = 2.4$, LMEM, max. random effects): NP sentences were read more slowly than PP sentences when the filler was animate, but not when the filler was inanimate. There were no differences among conditions in regions before the critical subject. The slowdown in animate NP conditions persisted after the subject, but didn't retain significance.

Discussion. We found a subject filled-gap effect only for relativized animates, suggesting that inanimate relativized arguments engender weaker predictions for an empty RC subject than animates. This is compatible with several flavors of expectation-driven parsing [6,7]. Interestingly, when we estimated the conditional probability of a subject gap given an RC-initial adjunct, we found that such RCs were almost always SRCs (P = .96, n=644). The influence of animacy was slight, though significant: P(SRC|*anim*) = .99, n=195; P(SRC|*inanim*) = .94, n=449. It would seem that adjunct phrases should strongly promote an SRC analysis regardless of animacy, at odds with our finding *no* filled-gap effect for inanimates. Either this suggests limited sensitivity to the adjunct's effect on continuations, or it argues in favor of a simple syntactic licensing pressure [7,10]. Animacy, while often dubbed a 'semantic' property, does have effects in English morphosyntax and selection. Our results are straightforwardly reconciled with syntactic theories that promote alignment between grammatical hierarchies, like animacy, person, grammatical function, argument structure, etc. [11].

References. [1] Traxler et al. (2002). *JML, 47,* 69-90 [2] Mak et al. (2002). *JML, 47,* 50-68. [3] Betancort et al. (2009). *QJEP 10,* 1915-1929. [4] Fedorenko et al. (2011). *CUNY 24.* [5] Staub (2010). *Cogn. 116,* 71-86. [6] Hale (2006) *Cogn. Sci. 30,* 643-672. [7] Aoshima et al. (2004). *JML 51,* 23-54. [8] Stowe, L.A. (1986). *LCP, 1,* 227-245 [9] Lee, M.W. (2004). *JPR, 33,* 51-73. [10] Pritchett (1992). [11] Aissen, J. (1999). *NLLT, 17,* 673-711.

Grammatically-guided resolution of filler-gap dependencies: An investigation of Chinese multiple dependencies

Jie Liu, Robert Fiorentino & Alison Gabriele (University of Kansas)
jieliu@ku.edu

Previous studies have provided evidence that the parser avoids positing gaps in grammatically unlicensed positions such as islands, suggesting that the grammar constrains the construction of filler-gap dependencies (e.g., Stowe, 1986). However, such evidence does not argue against an alternative interpretation positing that that islands are simply processing bottlenecks and thus the parser is unable to establish a filler-gap dependency within these domains (e.g., Hofmeister & Sag, 2010). To this end, Wagers and Phillips (2009) provide evidence that grammatical knowledge is not only used to avoid establishing wh-dependencies when prohibited, but is also used in motivating the active search for gap positions when required by the grammar. Building on Wagers and Phillips (2009), the current study examines whether gap-filling is predictive and grammatically-guided in topicalized sentences in Chinese. In a self-paced reading study (n=40), we compared the processing of sentences with coordinated verb phrases (1) to that of sentences with post-verbal adjunct phrases (2). Coordinated VPs require across the board extraction (*Who does John know _ and love _/*Mary?*) while extraction from post-verbal adjunct phrases is optional (*What did John eat _ after washing _/the dishes?*).

1. Coordinated VP:
   Nage wanhui / Nage ren Xiaowang **zhidao** t erqie yiqian **canjia-le** liang ci t, …
   That-CL party/#That-CL person Xiaowang know and before participate two times
   'That party/#that person Xiaowang knows and participated in twice before, …..'
2. Adjunct:
   Nage wanhui / Nage ren Xiaowang **zhidao** t yinwei yiqian **canjia-le** liang ci …
   That-CL party/That-CL person Xiaowang know because before participate two times
   'That party/#that person Xiaowang knows because he participated in twice before…..'

Thus, in (1), if the resolution of filler-gap dependencies is constrained by the grammar, the parser will actively search for a second gap after positing a gap in the first coordinated VP. We probed for this effect by manipulating the plausibility of the topicalized subject as an argument of the second verb *canjia-le (participate)*, with *party* as a plausible argument and *person* as an implausible argument. In contrast, in (2), this continued gap search is not required to satisfy grammatical constraints, although extraction from the second verb within the adjunct 'because' clause is possible in Chinese. If gap-positing is constrained by grammar, we expect a plausibility effect in (1) but not (2). Indeed, this pattern emerged in the spillover region directly following the second verb (and before the bottom-up information that there is a missing constituent), resulting in a Plausibility X Sentence Type interaction [$F(1,35)=6.99$, $p<.05$]. Follow-up *t*-tests revealed a significant reading time slowdown for the Implausible Coordinated VP as compared to its Plausible counterpart ($p<.001$), while no such slowdown was observed in the Adjunct conditions ($p>.1$). While this interaction was also present at the second verb itself (*canjia-le*) [$F(1,34)=5.89$, $p<.05$], follow-up *t*-tests revealed that the interaction emerged only due to a marginal slowdown for the Plausible Adjunct sentences ($p<.1$).

These results demonstrate, in line with Wagers and Phillips (2009), that grammatical knowledge actively guides the incremental resolution of filler-gap dependencies. These results also converge with those of previous studies (Huang & Kaiser, 2008) providing evidence for the establishment of filler-gap dependencies in topicalized structures in Chinese.

References

Hofmeister, P. & Sag, I. (2010). *Language, 86*, 366-415.
Huang, Y-C. & Kaiser, E. (2008). *Proc. of the 20th North American Conference on Chinese Linguistics, 2*, 927-941.
Stowe, L. (1986). *Language and Cognitive Processes, 1,* 227-245.
Wagers, M. W. & Phillips, C. (2009). *Journal of Linguistics, 45*, 395-433.

Increased linguistic focus leads to increased reading times

Matthew W. Lowder & Peter C. Gordon (University of North Carolina at Chapel Hill)
matt.lowder@unc.edu

Information that is linguistically focused in a sentence (by virtue of prosodic cues, discourse context, or sentence structure) enjoys a number of cognitive benefits relative to information that is not focused. A large literature demonstrates that linguistic focus attracts attention, improves anomaly detection, enhances memory, and improves the speed and accuracy of resolving anaphoric expressions. Despite these benefits, there is no clear consensus regarding how linguistic focus influences online sentence processing. Whereas some eye-tracking experiments suggest that reading times are longer for focused versus nonfocused information (e.g., Birch & Rayner, 1997), other experiments have provided evidence for the opposite pattern (e.g., Birch & Rayner, 2010). The current eye-tracking experiment improves on past experiments in two ways. First, it is important to note that previous experiments that have manipulated linguistic focus via syntactic structure have not adequately controlled for possible oculomotor differences between conditions. For example, Birch and Rayner (2010) analyzed reading times on a target word (e.g., *landlady*) that was either focused (e.g., *It was the landlady who…*) or neutral (e.g., *The landlady…*); however, the target word in the focused condition always appeared further into the sentence than the target word in the neutral condition. Second, previous experiments have not examined the effects of both linguistic focus and linguistic defocus relative to a neutral condition. Thus, the goal of the this experiment was to determine how reading-time patterns on a target word differ as the structure of the sentence makes the word increasingly more focused while ensuring that our results could not be explained on the basis of differences in oculomotor factors across conditions.

Participants read sentences like those presented in (1), where a target word (e.g., memo) could be defocused (1a), neutral (1b), or focused (1c). Figure 1 displays mean gaze durations and regression-path durations on the target word across the three conditions. Linear trend analyses showed that reading times increased as the target word became increasingly more focused (gaze duration: $F_1 = 5.61$, $p <$. 03; $F_2 = 3.52$, $p = .06$; regression-path duration: $F_1 = 6.56$, $p < .02$; $F_2 = 6.08$, $p < .02$).

Sentence structure acts as a powerful linguistic cue, indicating which constituents are more important to focus on than others. Readers are particularly sensitive to these cues, spending more time processing a word as its degree of focus increases. We believe these longer reading times reflect more complete encoding, which explains why readers have better memory for focused versus defocused information.

1a. It was the secretary that typed the official **memo** about the new policy. (Defocused)
1b. Yesterday the secretary typed the official **memo** about the new policy. (Neutral)
1c. What the secretary typed was the official **memo** about the new policy. (Focused)

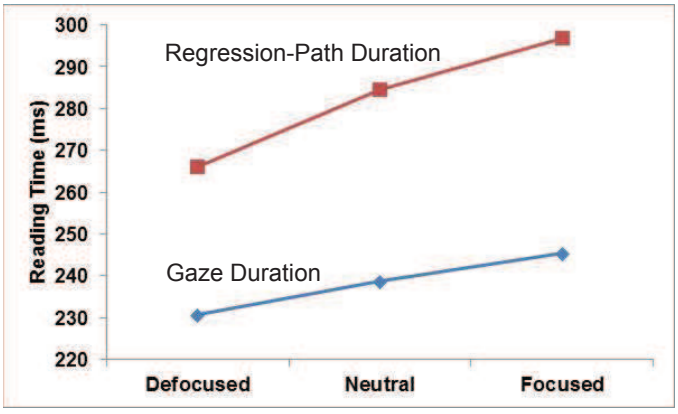


Figure 1

Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory & Cognition, 25*, 653-660.

Birch, S., & Rayner, K. (2010). Effects of syntactic prominence on eye movements during reading. *Memory & Cognition, 38*, 740-752.

Intrinsic differences in the processing of singular and plural pronouns

Mahayana C. Godoy (Campinas State University) & Cíntia Antão (Federal University of Minas Gerais)
mahayanag@gmail.com

Work on pronoun resolution has often assumed that these expressions elicit a set of antecedents as soon as they are read. However, the use of plural pronouns seems to defy such an incremental view, since their reference is often established later in the discourse. Consider the following text, whose Portuguese version with pronoun "eles"(plural/masculine) was used in a self-paced reading experiment: "Alice used to eat beef everyday, and **they** usually prepared a delicious New York strip **at the steakhouse**". In this text, a referent for the plural pronoun "they" is inferred through the locative "at the steakhouse", but it is not until the end of the sentence that the reader has this information. These observations raise the question of whether plural pronouns really need an antecedent by the time they are read, an assumption made by most of the studies about the processing of plural pronouns [e.g., 1,2].

Previous research had already found that plural pronouns with inferred antecedents do not impose extra processing cost, but this study also states that plural pronouns need an identifiable antecedent, be it explicit or inferred [3]. In a self-paced reading task, we tested whether a plural pronoun really needs an antecedent. 48 subjects read 42 texts similar to the six conditions presented in (1). The "no-antecedent" condition was controlled to build a discourse representation through which no antecedent could be recovered. The focus of the sentence was always on a female character. Inferred contexts had a locative at the beginning of the sentence that could trigger an unspecific referent for the plural pronoun (e.g., "at the steakhouse" > "the steakhouse's staff)". Co-referential contexts served as a neutral baseline to compare reading times. We measured reading times for the sentences in italic.

(1)	a) No-antecedent	Alice used to eat beef everyday, *and **they/he** usually prepared a delicious New York strip* **at the steakhouse.**
	b) Inferred antecedent	**At the steakhouse**, Alice used to eat beef everyday, and ***they/he*** *usually prepared a delicious New York strip* with onions.
	c) Co-referential antecedent	**Robert and Paul/Robert** used to eat beef everyday, and ***they/he*** *usually prepared a delicious New York strip* with onions.

Our analysis showed an antecedent effect on the processing of singular pronouns. For these sentences, the adverb placed after the pronoun to control for spill-over effects showed the same reading times for conditions (a) and (b), but these times were greater than the one observed for condition (c) ($p<0,001$ for (a) vs (c); $p=0,002$ for (b) vs (c)*). This effect was not found for plural pronouns, meaning that the processing cost for "they" was the same across all conditions.

From the perspective of pronoun processing, we can conclude that the results described above do not fit the presupposition that singular and plural pronouns processing follow the same resolution strategy. Because plural pronouns may depend on information presented later in the discourse, its resolution may be delayed when no referent is found in previous context. Singular pronouns, on the other hand, trigger an immediate search for an antecedent, causing extra processing cost when it is not found. Furthermore, because our data suggest that plural pronouns may continue unresolved without causing extra processing cost, these results may be also interpreted as an evidence that language processing, in specific situations, may rely on superficial representation of discourse structure [4].

*Data resulted in a non-normal distribution; therefore, we used the Mann-Whitney test.

References

[1] Oakhill et al.(1992) *Lang. and Cog. Proc.*
[2]Gelormini&Almor(2013) *Journal of Psyc Res.*
[3] Filik et al., (2008) Journal of Cog. Sc.
[4] Sanford&Sturt (2002) Trends in Cog. Sc.

The Middle matters the most: The effect of phonological similarity on referring forms

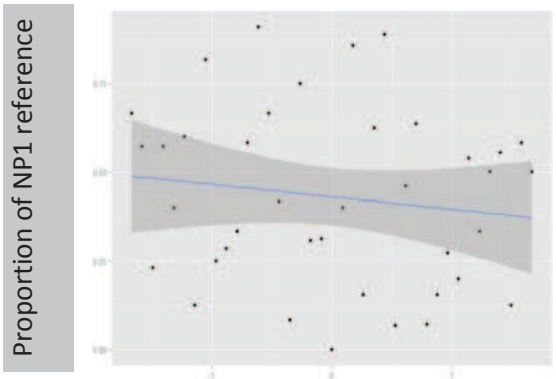
Hossein Karimi & Fernanda Ferreira (University of South Carolina)
karimi@email.sc.edu

Speakers' choice of referring expression is influenced by a range of linguistic factors; in this study, we investigated the potential effect of phonological similarity between two mentioned entities. We used 40 pairs of proper names differing in phonological similarity on a continuous scale from 1 to 7 (as rated by 12 native English speakers). Frequency, length, familiarity and gender of the names were matched. Sixteen participants provided written continuations to given sentences which contained names varying in phonological similarity as in (1). The use of pronouns (versus repeated names) to refer to the entities in the given sentence was measured.

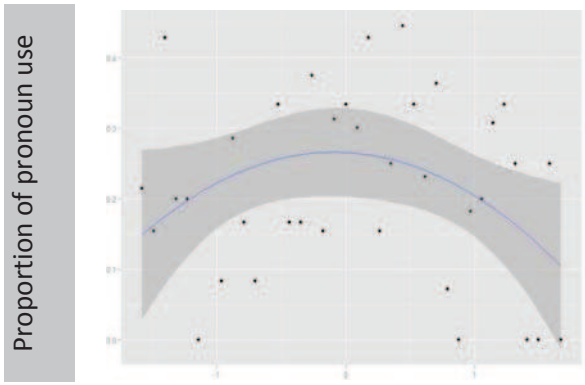
(1)

a. Frank found Joe on Facebook. (Mean phonological similarity = 1.1)
b. Albert recued Arthur in the mountains. (Mean phonological similarity = 3.4)
c. Donald bumped into Ronald in the street. (Mean phonological similarity = 6.2)

Using logistic polynomial regression models, we analyzed both choice of referent (i.e., which character the participants referred to in their continuations) as well as form of referring expression (i.e., the linguistic form participants used to express that choice). Analyses of referent choice revealed no relationship between phonological similarity and choice of referent (Figure 1). However, the results for choice of referring expression revealed a quadratic relationship between phonological similarity and pronoun use such that there was less pronoun use for highly dissimilar (Frank and Joe) and highly similar (Donald and Ronald) names. However, there was more pronoun use for names that fell in the middle of the similarity continuum (e.g., Albert and Arthur; Figure 2).



Phonological similarity
Figure 1



Phonological similarity
Figure 2

The results suggest that phonological interference affects pronoun use, consistent with theories of language production that assume feedback from phonology to syntax (e.g., Dell, 1986). In addition, the results are broadly consistent with studies showing that phonological similarity as measured by “phonological neighborhood density” causes linguistic reduction (Gahl, Yao & Johnson, 2012) and production difficulty (Jaeger, Furth, & Hilliard, 2012). However, our data also suggest that the relationship between phonological similarity and linguistic form is not linear; instead, extensive similarity and extensive dissimilarity facilitate production of the associated word (the proper name, in this case), but moderate similarity causes interference and thus gives rise to production of alternative forms such as pronouns.

Noun-phrase-internal structural priming in a picture-description task

Amy DiBattista & Neal J. Pearlmutter (Northeastern University)
dibattista.a@husky.neu.edu

In structural priming, the syntactic structure of a produced or comprehended utterance is reused in a subsequent utterance; and the processing of an utterance facilitates the processing of a subsequent one with the same structure (Pickering & Branigan, 1999). In the priming of noun phrase structure, speakers reuse structures (e.g., Det–N–RC in *the square that's red*) after hearing them from an interlocutor (Cleland & Pickering, 2003); and picture description latencies are speeded when immediately preceded by a description with the same structure (Wheeldon & Smith, 2003). The current work investigated phrase-internal structural priming, or structural reuse from one noun phrase (NP) (e.g., *the green spot*) to the next (e.g., *the blue apple*) in a complex noun phrase (CNP; e.g., *the green spot on the blue apple*). The likelihood of a structure in the NP produced second (NP2) was predicted to increase given the use of that structure in the NP produced first (NP1); and CNPs with structurally homogeneous NPs (e.g., *the green spot on the blue apple*) were predicted to show facilitation in the form of lower error rates compared to heterogeneous CNPs (e.g., *the green spot on the apple that's blue*).

171 participants produced noun phrase–prepositional phrase descriptions for 32 pictures, each with two components, in four color schemes: one with two solid colors, one with two checkered colors, and two versions with one solid and one checkered color. Participants were instructed to describe solid components with a prenominal adjective (Det–Adj–N, e.g., *the green spot*) and checkered components with an adjective in a relative clause (Det–N–RC, e.g., *the apple that's blue*). Responses were homogeneous (e.g., *the green spot on the blue apple*, *the spot that's green on the apple that's blue*) or heterogeneous (e.g., *the green spot on the apple that's blue*, *the spot that's green on the blue apple*) in structure. Participants first previewed grayscale versions of the pictures and their noun labels. Then, during testing, each picture appeared in a color scheme, and the participant described it using the Ns, a given preposition, and the Adjs as instructed.

Structure error rates, or the percentage of responses in which at least one NP was produced in a structure opposite what was instructed, were analyzed. By-participant and by-item empirical-logit linear regressions were conducted on structure error rates (fixed effects: intended NP1 structure, intended NP2 structure, and their interaction) and on rates of homogeneous versus heterogeneous responses (fixed effects: homogeneity of the intended response, produced NP1 or NP2 structure, and their interactions). Responses were more likely to be homogeneous than heterogeneous overall; and structure errors were less likely for homogeneous responses. Error responses had homogeneous NP2s more often than heterogeneous NP2s (e.g., if Det–Adj–N was produced in NP1, Det–Adj–N was more likely than Det–N–RC in NP2). Error responses had homogeneous NP1s more often than heterogeneous NP1s given a Det–N–RC in NP2. This bias was not present given a Det–Adj–N in NP2.

The reuse of NP1 structures in NP2s is a clear example of phrase-internal structural priming. Structural priming in production therefore can occur from one part of a CNP to the next part, as that CNP is planned. This may facilitate production by eliminating the need to build NP2's structure from scratch. The effect is strong enough to override explicit instructions to produce heterogeneous structures.The priming from NP2 to NP1 suggests that NP2 and NP1 may overlap in planning, allowing Det–N–RC structures (perhaps due to stronger priming from unpreferred structures [summarized in Ferreira & Bock, 2006]) in NP2 to be reused in NP1.

References

Cleland, A. A., & Pickering, M. J. (2003). *Journal of Memory and Language*, *49*, 214–230.
Ferreira, V. S., & Bock, K. (2006). *Language and Cognitive Processes*, *21*, 1011–1029.
Pickering, M. J., & Branigan, H. P. (1999). *Trends in Cognitive Science*, *3*, 136–141.
Wheeldon, L., & Smith, M. (2003). *Language and Cognitive Processes*, *18*, 431–442.

Decoding the semantics of words in sentences from neural activity

Alona Fyshe*, Erika Laing, Nicole Rafidi, Kai-min Chang, Tom Mitchell (Carnegie Mellon University)
*afyshe@cs.cmu.edu

The neural representation of single word semantics has been relatively well studied, but the effect of sentence context on this representation is largely underexplored. We used machine learning techniques to study how sentence processing interacts with individual word semantics. We examined the accuracy of decoding individual words from neural activity when words were presented in isolation, and in sentences of varying semantic predictability. Prior work (Murphy 2012, Mitchell 2008, Fyshe 2013) has successfully decoded corpus-derived features of stimulus words from brain activity; here we apply the technique to sentence reading.

We collected magnetoencephalography (MEG) recordings of people reading 45 words in isolation (8 repetitions each, randomized), followed by 30 simple noun-verb-noun sentences built from those 45 words (8 repetitions each, randomized). Half the sentences were *predictable,* with the verb highly predicted by the first noun (e.g. Pilot flies plane.), and half were *less predictable*, with the first noun and verb semantically consistent, but not strongly related (e.g. Pilot scrubs teeth). All words in all conditions were presented for 800ms, with at least 200ms of rest following. Participant engagement was ensured by probing comprehension with semantic questions after 10% of trials, in all conditions. We processed the recordings with tSSS, downsampled to 200Hz, band-pass filtered from .1 to 50Hz, and removed ocular and cardiac artifacts with SSP. Words were parsed into epochs of -200 to 800ms and baseline corrected, and further analysis used data from stimulus onset to 750ms.

We created an independent measure of word semantics by extracting word usage patterns from a 16 billion word corpus. We used the Malt parser to dependency-parse the corpus and compiled dependency statistics about the relationships between words in sentences. We used singular value decomposition (SVD) to compress the resulting matrix to 1000 dimensions, which forms a “semantic vector” for each word. Across our subject population, peak decodability was reached using the top 30-50 SVD dimensions.

We trained independent L2-regularized regressors to predict each of the semantic vector dimensions as a function of the MEG data. During leave-one-out cross-validation, the full set of 45 words were ranked by their cosine distance to the semantic vector predicted for the held out brain image. We report rank accuracy: the percentage of words amongst the 45 words with cosine distance larger than the distance to the correct word. Chance performance for rank accuracy is 50; [45.6 - 54.4] forms a 95% confidence interval.

We found rank accuracy in all conditions to be much higher than chance. Accuracy is higher for words in context (81% across all words in both sentence conditions) than for words in isolation (69% across all words), implying that sentence context can amplify the semantics of single words and lead to heightened decodability. We found no significant difference in accuracy between sentence conditions, suggesting that either predictability does not affect decodability, or that the differences in predictability we explored are not sufficient to elicit a difference. These preliminary results are an intriguing first step towards understanding the incremental semantic processes underlying sentence processing. To extend this work, we plan to examine the decodability of words over the entire sentence, rather than time-locked to each word presentation. This could reveal the development of the subject’s internal situation model, as well as demonstrate prediction and reactivation in sentence processing.

Research supported in part by grants from NICHD (5R01HD075328-02), and from IARPA (FA865013C7360).

Fyshe, A., et al. Documents and Dependencies: an Exploration of Vector Space Models for Semantic Composition: International Conference on Computational Natural Language Learning (CoNLL 2013), Sofia, Bulgaria.

Mitchell, T. M., et al.(2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191-1195.

Murphy, B., Talukdar, P., & Mitchell, T. (2012, June). Selecting corpus-semantic models for neurolinguistic decoding. *S*EM* (pp. 114-123). Association for Computational Linguistics.

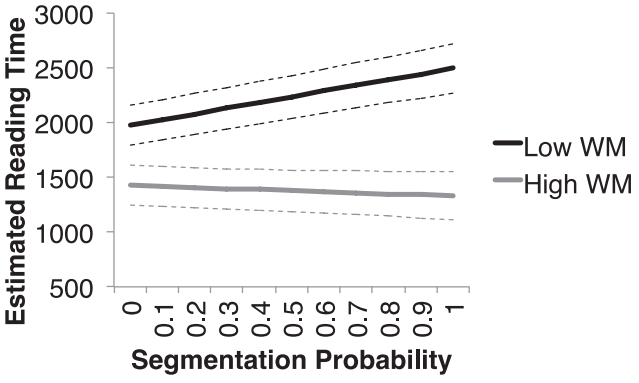
Eye movements reflect the cognitive reality--and costs--of event structure during reading

Benjamin Swets & Christopher A. Kurby (Grand Valley State University)
swetsb@gvsu.edu

When readers perceive an event shift to occur while reading narrative texts, they update their representations of "what is happening now" (Kurby & Zacks, 2008). This shift in representations from one event model to the next implies the existence of boundaries between events in narratives. For example, readers will slow down their reading rate during situational changes, and they lose availability of previous information after a shift in the situation (Zwaan, 1996). Other research has found that when presented with a narrative text presented one clause at a time, reading rate slows down at clauses that had previously been labeled event boundaries, over and above effects accounted for by other situational changes (e.g., a change of character, or goals) (Zacks et al., 2009). The purpose of the present experiments was to use eye movements to explore the moment-to-moment processing consequences of event cognition during more naturalistic reading of larger chunks of text (i.e., more than one clause at a time).

In Experiment 1 (N = 34), participants read a narrative account of a boy's day while their eye-movements were recorded. In Experiment 2 (N = 84), we replicated the procedures of Experiment 1 but also assessed working memory capacity. Based on previous work on event cognition (Kurby & Zacks, 2008), we hypothesized that reading times (e.g., first run reading time, first fixation duration, total reading time) should be slower for event boundaries than non-boundaries. Readers should also be more likely to make regressive eye-movements into event boundaries. In addition, if readers update their event models at event boundaries and lose accessibility of previous event information at such points, readers will be less likely to make regressions out of event boundaries than non-boundaries. Finally, if event updating incurs a working memory load, individuals with lower working memory capacities should show stronger effects of event boundaries on reading behavior.

We conducted mixed effect models with a continuous event segmentation likelihood variable as the main predictor. We also included terminal punctuation, non-terminal punctuation, and number of characters as covariates. In both experiments we found that reading time was significantly slower at event boundaries clauses, as measured by total fixation time, first run fixation time, and even first fixation time. We also found that readers were significantly more likely to make regressive eye movements into event boundary clauses, but were less likely to regress out of an event boundary clause. In Experiment 2, we also found that working memory capacity interacted with event boundary status for total reading time, such that low working memory readers slowed down at event boundaries, but high working memory readers did not (see figure). It appears that eye movements strongly reflect the cognitive reality of event structure during reading, and also that tracking this event structure indeed carries a working memory load.



References

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, *12*, 72–79.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, *138*, 307–327.

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1196–1207.

Implicit Agents in Short Passives and Remote Control of Reason Clauses

Michael McCourt, Aleksandra Fazlipour, Ellen Lau and Alexander Williams (U. of Maryland)
mmccour2@umd.edu

In understanding speakers we make use of implicit meaning, corresponding to no audible linguistic expression. In some cases, linguistic patterns suggest that this meaning corresponds to a silent grammatical constituent or feature, while in other cases the source of this meaning appears to be non-linguistic. However, previous work has noted that when linguistic patterns cannot distinguish whether implicit meaning is grammatical or pragmatic, processing data can be informative (Mauner et al., 1995; 2000). The goal of the self-paced reading time study presented here was to begin to clarify the status of the *implicit agent* of short passives, (1).

(1) Two outfielders were traded away.

Short passives render implicit the role bound by the subject in the active. Yet this *implicit agent* behaves in one way like its explicit counterpart. It can *control* an infinitival reason clause, something not generally possible without an explicit antecedent (Roeper, 1987). In other words, the understood acquirer of a pitcher in (2) can be the implied agent of trading. Furthermore, processing studies observe no additional cost for processing the infinitival reason clause when it is controlled by an implicit vs. explicit agent (2-3; Mauner et al. 1995, 2000).

(2) Two outfielders were traded away to acquire a better pitcher.

(3) Some team traded away two outfielders to acquire a better pitcher.

These results have been taken to support the *Grammatical Theory*: at some level of syntax, short passives are identical to actives, and control is stated at that level as a grammatical dependency between two clauses. The logical subject of the reason clause is treated as a bound variable. Yet this theory of *Local Control* cannot account for *Remote Control* in (4,5).

(4) Two outfielders were traded away. The goal was to acquire a better pitcher.

(5) Some team traded away two outfielders. The goal was to acquire a better pitcher.

A fortiori Remote Control does not involve a *grammatical* dependency, since these do not cross sentences. If Remote Control is broadly acceptable to speakers, we need a *Discourse Theory*, on which the logical subject of the reason clause behaves like an anaphoric definite description. In sum, the Grammatical Theory requires that Local and Remote Control involve two different mechanisms, while the Discourse Theory could (but need not) apply similarly in both.

As a first step towards examining the mechanisms supporting Local vs. Remote Control, we measured word-by-word self-paced reading in a 2x2 design (Active/Passive x Local/Remote), as in (2-5). Main regions of interest were “to” and the verb, where the control relation is resolved. Two clear results are observed. First, no slowdown is observed in Remote relative to Local, suggesting that control is no more difficult in the Remote case. Indeed RTs at “to” are faster in the Remote passive ($p<.05$, by subjects and items), though this may be an effect of first interpreting “the goal”, a cue not present in Local. Second, within the Remote conditions there is no difference between Passive and Active ($p>.2$, by subjects and items), hence no effect of whether or not an explicit controller is available. This suggests that the observation of no processing cost for implicit control in Local cases (Mauner et al. 1995, 2000) cannot be taken as positive evidence for the Grammatical Theory, since the same is observed in the Remote case where control cannot be grammatically mediated. Also, unlike Mauner et al., we do find some evidence for a slowdown for Local Passive relative to Active at the verb ($p<.05$ by subjects).

This first study of Remote Control neither supports nor refutes the Grammatical Theory of Local Control; nor does it demonstrate whether Control is implemented by the same mechanism in both Local and Remote cases. It does, however, confirm the challenge to the Grammatical Theory posed by Remote Control, and provides the first measures of its online interpretation, suggesting new questions concerning both the representation of short passives and the processing of control.

References: Mauner, G. et al. 1995, *JML* 34. Mauner, G. & J.-P. Koenig 2000. *JML* 43. Roeper, T. 1987. *LI* 18, 267-310. Williams, A. 2013. *Argument Structure*, ms. CUP.

Unconscious lexical semantic activation of form neighbors

Dane Bell, Kenneth I. Forster, and Thomas G. Bever (University of Arizona)
dane@email.arizona.edu

This work investigated early, unconscious lexical semantic processing and its neural and behavioral differences in right-handed individuals with and without left-handed relatives.

In a semantic categorization task, real-word masked primes that are form-similar to members of the same category as the target reduce response time relative to targets with primes that are form-similar to members of an opposing category to the target (Forster & Bell, 2013). For example, since *pinch* is similar to *peach*, a trial such as *pinch-APPLE* in which *pinch* is very briefly presented would be responded to faster than one such as *hazard-APPLE* if the category is FRUIT. (Just as *pinch* is close to *peach, hazard* is close to *lizard*.) Such an advantage is possible only if *pinch* is capable of activating the meaning of *peach*, which requires that word form recognition overlaps in time with semantic activation. This suggests that some semantic content of a word is available early, before the form of the word is fully recognized; this supports a cascaded activation account (and contradicts a staged activation account).

In order to study the time-course and neural organization of this semantic activation, EEGs were collected from participants in the same paradigm. Right-handed subjects were categorized as with (FS+) or without (FS–) family left handers, because previous work has shown neurological and behavioral differences between these groups (Hancock & Bever, 2013; Sammler et al., 2012). Behaviorally, right handed FS+ individuals are better at single-word lexical access, whereas FS– individuals are better at syntactic processing (Bever et al. 1987).

Event-related potential (ERP) analyses were triggered by the onset of the target noun. Across all participants, incongruent (*hazard-APPLE*) in-category target trials correlated with sustained increased positivity starting at 200 ms, relative to congruent primes (*pinch-APPLE*), in the midline and right frontal and temporal lobes. Most of that right hemisphere congruence effect is due to FS+ subjects. Furthermore, incongruent out-of-category target trials (*pinch-LION* if the category is FRUIT) correlated with *decreased* positivity shortly after the target onset, relative to congruent out-of-category target trials (*hazard-LION*), in the *left* temporal and occipital lobes. Most of that left hemisphere congruence effect is also due to FS+ subjects.

The results support the notion of cascaded activation at an early stage of neurological processing: the congruence effects require that some aspects of lexical semantics are activated before the whole word form is recognized. Furthermore, the superiority of the in-category congruence effect in the right hemisphere areas is consistent with previous studies of “coarse” semantic categorization (see Lindell 2006). Finally, virtually all of the right lateralization congruent effect for in-category target trials and the left lateralization effect for out-of-category target trials is due to FS+ subjects. Thus, the results also contribute new evidence that familial sinistrality is an index for widespread differences in the neural organization of language. Since approximately 40% of the population are right-handed FS+, it is now critical to track familial handedness in neurological and behavioral language studies.

References

Bever, T. G., Carrithers, C., Cowart, W., & Townsend, D. J. (1989). Language processing and familial handedness. *From reading to neurons: Issues in the biology of language and cognition*, 331-357.

Forster, K. I., & Bell, D. (2013). Is cascaded semantic activation semantic? Is it cascaded? Paper presented at the 2013 Psychonomic Society Annual Meeting, Toronto, Canada.

Hancock, R., & Bever, T. G. (2013). Genetic factors and normal variation in the organization of language. *Biolinguistics*, *7*, 75-95.

Lindell, A. K. (2006). In your right mind: Right hemisphere contributions to language processing and production. *Neuropsychology review, 16(3),* 131-148.

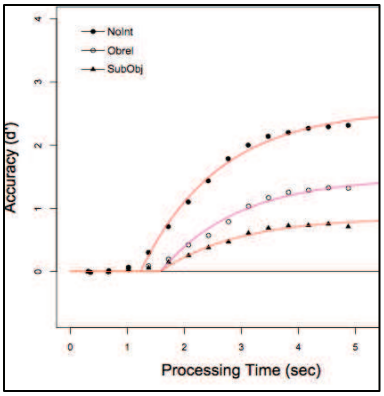
Sammler, D., Hancock, R., Bianco, R., Friederici, A. D. and Bever, T. G. (2012). Genetic factors in the cerebral asymmetries for language and music. Poster presented at the Neurobiology of Language Conference, San Sebastian, Spain.

A speed-accuracy tradeoff study of poor readers' memory mechanisms

Clinton L. Johns, Julie A. Van Dyke, and Dave Kush (Haskins Laboratories)
johns@haskins.yale.edu

People vary widely in their ability to comprehend sentences. Many accounts of sentence comprehension attribute this to intrinsic individual differences in working memory capacity (WMC): low WMC leads to comprehension failure because readers are unable to actively maintain necessary information (e.g., Just & Carpenter, 1992). However, an alternative account suggests that active memory is extremely limited *even in skilled comprehenders*, and that a content-addressable memory retrieval mechanism restores information into active memory when it is needed (e.g., McElree, Foraker, & Dyer, 2003). On this account, comprehension fails when retrieval fails. The current study seeks to evaluate the possibility that poor comprehension arises from a less efficient retrieval mechanism as compared to that used by skilled readers.

We replicated a definitive experiment (McElree et al., 2003, Exp 2) with a community-based sample of non-college bound young adults (n=20; ages 16-24), who displayed a wide range of linguistic ability. As in McElree et al., we used the speed-accuracy tradeoff method (SAT), which provides a measure of response accuracy as it changes over time that is not confounded by participant-specific response thresholds. Three innovations to the original study were introduced: 1) we used the multiple-response SAT method, which requires fewer testing sessions and is less demanding on participants; 2) stimuli were presented auditorially, so as to examine the retrieval mechanism independent of potential effects of reading difficulty; 3) we administered a battery of individual difference (ID) measures to index WMC, vocabulary, oral language ability, reasoning, phonological and spatial awareness, and IQ.



The study design was simplified from the original, and included three syntactic constructions (1-3). The study assessed whether the speed of integrating the final verb with its subject was affected by the amount of intervening material. Capacity-based accounts predict that more information between subject and verb will increase retrieval speed because more information must be actively maintained while searching for the subject. In contrast, a content-addressable retrieval mechanism predicts that retrieval speed will not vary, except that the immediately adjacent condition (1) will be faster than (2) and (3), which will not differ; and accuracy will vary by condition due to increased interference, which affects the probability (but not speed) of retrieving the correct subject (Van Dyke & McElree, 2011).

(1) No intervening material: The book ripped.
(2) Object relative clause: The book that the editor admired ripped.
(3) Object + subject relative: The book that the editor who quit the journal admired ripped.

Preliminary analyses using mixed models assessed speed, accuracy, and ID measures (each modeled separately, as in Van Dyke, Johns, & Kukona, in press). Consistent with content-addressable retrieval, the amount of intervening material did not affect overall speed of retrieval. Moreover, no ID measures interacted with speed. Main effects of oral and reading comprehension were observed, such that low ability resulted in slower retrieval speed overall, but these slow-downs were unrelated to the amount of interpolated material. As expected, accuracy varied across conditions, but there were no interactions with skill. A main effect of phonological awareness was observed, such that lower skilled readers had lower overall accuracy. We conclude that poor readers, although slow, still use content-addressable retrieval. Thus, poor comprehension is not a result of a qualitatively different retrieval process.

References: Just, M. A., & Carpenter, P. A. (1992). *Psychol Rev*, *99*, 122-149; McElree, B., Foraker, S., & Dyer, L. (2003). *J Mem Lang*, *48*(1), 67-91; Van Dyke, J. A., & McElree, B. (2011). *J Mem Lang*, *65*(3), 247-263.

Effects of integration in eye tracking

William Schuler & Marten van Schijndel (The Ohio State University)
schuler@ling.osu.edu

Probabilistic parallel processing models (Hale, 2001; Levy, 2008) provide an attractive explanation of frequency effects in reading, but usually have little to say about effects of working memory operations like encoding and integration. The present work evaluates predictions of a model which not only measures n-gram and incremental PCFG probabilities of parallel hypotheses (surprisal), but also computes connected components of partial syntactic dependency structures in each hypothesis, and measures the proportion of parallel hypotheses in which disjoint connected components of syntactic dependency structure must be recalled and composed. In particular, this work attempts to explain an apparent discrepancy between previous experiments on the same data set (the Dundee corpus; Kennedy et al., 2003), one of which shows a significant positive effect on first-pass fixation durations for integration operations weighted by dependency length using a single hypothesized parse (Baumann, 2012), and one of which shows no positive effect for integration cost (and instead a significant negative effect) using a measure of integration weighted by the proportion of parallel incremental hypotheses involving integration (van Schijndel et al., 2013). Is positive integration cost a casualty of parallel processing models?

To answer this question, the parallel integration measure of van Schijndel et al. (2013) was evaluated with the same basket of predictors as Baumann (2012): word position, word length, unigram and bigram frequency, and all two-way interactions, using linear mixed-effects regression. Results show a numerically but not significantly positive effect for integration cost ($.19$ms± 1.2ms, $p{=}.88$). Then random slopes for word position, word length, unigram frequency, and integration cost were added (slopes for bigram frequency did not converge), with the result that the sign for the effect of integration cost turned numerically but not significantly negative ($-.20$ms± 1.8ms, $p{=}.91$). Then an indicator was added showing whether the immediately previous fixation was to the immediately previous word. With random slopes for this and the above predictors, this increased the strength of the negative effect for integration cost to be statistically significant (-3.4ms± 1.6ms, $p{<}.05$). Finally, PCFG surprisal was added as a predictor as defined in van Schijndel et al. (2013), with results showing no positive effect for integration, and instead a strongly significant negative effect for integration (-6.3ms± 1.9ms, $p{<}.001$), with random slopes for this and the above predictors.

These results suggest that loss of positive effect for integration cost is not a consequence of adoption of a parallel processing model, but rather of a combination of predictors relating to fixation history and PCFG surprisal, and the use of random slopes for predictors. Since these predictors are independently motivated and increase model fit significantly and substantially, these results cast some doubt on the existence of a broad positive effect for integration cost on fixation durations in eye-tracking data, and may lend urgency to efforts to eliminate possible confounds (perhaps related to grammar rule probabilities) in experiments on constructed stimuli (Bartek et al., 2011).

References

Bartek, B., Lewis, R. L., Vasishth, S., and Smith, M. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 37(5):1178–1198.

Baumann, P. (2012). The role of hierarchical structure in syntactic dependency integration. In *Twenty-Fifth Annual CUNY Conference on Human Sentence Processing*, page 91.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of NAACL*, pages 159–166, Pittsburgh, PA.

Kennedy, A., Pynte, J., and Hill, R. (2003). The Dundee corpus. In *Proceedings of the 12th European conference on eye movement*.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.

van Schijndel, M., Nguyen, L., and Schuler, W. (2013). An analysis of memory-based processing costs using incremental deep syntactic dependency parsing. In *Proceedings of CMCL 2013*.

Poster #3026

Eye-tracking and event-related potential effects of transposed letters during compound word comprehension: Do morphemes matter?

Mallory C. Stites, Kara D. Federmeier, & Kiel Christianson (University of Illinois)
Stites2@illinois.edu

The current study investigates the online processing consequences of encountering compound words with transposed letters (TLs) in a sentence context, to determine if TLs that cross morpheme boundaries are more disruptive to reading than those within a single morpheme. Previous eye-tracking work has demonstrated that reading text with TLs produces general reading time costs relative to correctly spelled words (Rayner, White, Johnson, & Liversedge, 2006). There is an ongoing debate in the literature as to whether multimorphemic words are recognized via morphological decomposition in English, and behavioral studies employing across-morpheme letter transpositions as a test case have found mixed evidence as to whether they are more disruptive to processing than within-morpheme TLs (Christianson, Johnson, & Rayner, 2005) or not (Rueckel & Rimzhim, 2011). To address this disparity, two separate experiments were conducted, one employing eye-tracking during natural reading, and the other using event-related potentials (ERPs) during word-by-word sentence reading. In both experiments, subjects read weakly- to moderately-constraining sentences containing compound words in a sentence-medial position. The compounds could appear as correctly spelled (*cupcake*), or with TLs either across morpheme boundaries (*cucpake*) or within one morpheme (*cupacke*), which appeared equally in the first and second morpheme across items. In both studies, subjects read every compound word once, in one of the experimental conditions. In the eye-tracking study, reading times were longer for all TL words relative to correctly spelled words, with no difference between the two TL conditions. These findings suggest that readers do not obligatorily decompose compounds into their constituent morphemes during natural reading, as cross-morpheme letter transpositions would be more disruptive to this process than within-morpheme TLs.

Eye movements, however, as an end-state measure, might fail to show earlier effects that transposed letters may have on semantic analysis of the compound word. As such, we conducted a second study using ERPs to see if there are differential costs to early semantic analysis evident on the N400 component that would not ultimately show up on eye tracking costs. Findings showed no difference between the TL and correctly spelled compound words on the N400. Instead, the TL words elicited a late posterior positivity (from 600-900 ms) relative to correctly spelled words, a component often elicited by conflict resolution processes. Importantly, the amplitude and timing of the late positive component did not differ between the two TL conditions. The ERPs thus allow us to add detail to the eye-tracking results by targeting which aspect of processing the transposed letters affect—providing evidence that they do not differentially impact the reader's ability to extract meaning from the stimulus. This study provides evidence against the idea that compound words are obligatorily recognized via decomposition into their constituents, which would predict greater processing costs from across-morpheme TLs, especially on the N400. Rather, letter transpositions cause processing costs at a later stage of processing indexed by the late posterior positivity, perhaps in overcoming the conflict between the misspelled input and readers' existing representation of the word.

References

Christianson, K., Johnson, R. L., Rayner, K. (2005). Letter transpositions within and across morphemes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*(6), 1327-1339.
Rayner, K., White, S. J., Johnson, R. L., & Liversedge, S. P. (2006). Raeding Wrods with Jubmled Lettres: There Is a Cost. *Psychological Science, 17*(3), 192-193.
Rueckl, J. G. & Rimzhim, A. (2011). On the interaction of letter transpositions and morphemic boundaries. *Language and Cognitive Processes, 26*(4/5/6), 482-508.

On-line sentence reading in people with aphasia: Evidence from eye tracking

Gayle DeDe & Jessica Knilans (University of Arizona)
gdede@arizona.edu

According to the Lexical Bias Hypothesis (e.g., Gahl, 2002), people with aphasia (PWA) are sensitive to conflicts between sentence structure and the biases of the words in the sentence. It is unclear whether this hypothesis can be extended to include biases – or expectations – based on the relative frequency of different syntactic structures. Recent work suggests that both syntactic expectations and syntactic complexity affect how college students read sentences with object and subject relative clauses (Staub, 2010). With respect to syntactic expectations, Staub reported reading times for the second noun phrase, which is the first point at which an object relative structure can be detected. Go-past reading times were longer for the second noun phrase in object compared to subject relative clauses, suggesting participants were slowed by the violation of the expectation for the more frequent structure (i.e., the subject relative; cf. Roland et al., 2007). With respect to syntactic complexity, Staub (2010) reported longer reading times for the verb in object compared to subject relatives. Reading times for the verb were interpreted as evidence of processes involved in building a more complex structure.

The present study used eye tracking to examine how PWA read object and subject cleft sentences. Following Staub (2010), we predicted that PWA would show effects of structural frequency at the second noun phrase and effects of syntactic complexity at the verb. Previous studies have demonstrated that syntactic complexity affects how PWA process object and subject cleft sentences, but this is the first study to address this question using eye tracking during reading. This is also the first study to directly investigate whether structural frequency affects comprehension in PWA, as would be predicted by the Lexical Bias Hypothesis.

Nine PWA (4 Anomic, 4 Broca, 1 Conduction) and eight age-matched controls read object and subject cleft sentences while a camera recorded their eye movements. Sentences were constructed in pairs and varied only with respect to word order (examples 1 & 2). Nouns were matched for frequency and length across sentence pairs. Each sentence was followed by a comprehension question (e.g., *Did the baby entertain the father?*).

1. Object Cleft: It was the baby that **the father entertained** during the party last week.
2. Subject Cleft: It was the father that **entertained the baby** during the party last week.

In general, PWA read significantly more slowly and made more regressive eye movements than the controls. Here, we present go-past and rereading times from two critical regions (bolded in 1 & 2). At the second noun phrase (NP2), both groups showed longer go-past reading times in object compared to subject clefts, suggesting that they were sensitive to the violation of the expectation for the more frequent structure. The groups showed opposite patterns of rereading times for NP2: Controls had longer rereading times for object than subject clefts, whereas PWA had longer rereading times for subject than object clefts. For the verb, controls showed the predicted pattern of longer reading times in object versus subject clefts for all measures. Surprisingly, PWA's reading times were longer for subject than object clefts in early measures (e.g., go-past times). This pattern was reversed in later measures; PWA's rereading times were longer for the verb in object compared to subject clefts.

The reading times for NP2 suggest that the Lexical Bias Hypothesis can be extended to include at least some biases based on structural frequency. However, the eye-tracking data also revealed that PWA did not show normal on-line processing of subject and object clefts. PWA tended to read relatively quickly on the first pass through the sentence. On subsequent passes, PWA spent more time rereading the last word of the clause: the verb in object clefts and NP2 in subject clefts. These rereading times may reflect more time spent in clausal wrap-up processes, which would be consistent with the idea that lexical integration is impaired in PWA (cf. Thompson & Choy, 2009).

Why do interference effects arise in sentence processing?
A sampling theory of memory as optimal discrimination among noisy traces

Richard L. Lewis, Michael Shvartsman & Satinder Singh (University of Michigan)
rickl@umich.edu

Overview. We present a theory and computational model of short-term memory (STM) in sentence processing that (a) explains (rather than stipulates) slow-downs due to similarity of retrieval candidates (cf. Gordon et al 2002; Xiang et al 2013; Van Dyke, 2007); (b) includes a strategic retrieval threshold that can function in a broader theory of adaptive, "good-enough" processing, (c) provides the first computational account of speed-accuracy tradeoffs observed in response-deadline studies of reading (e.g. Van Dyke & McElree, 2011); and (d) yields a unified basis for understanding how sentence processing handles uncertainty in both perception and memory.

Theory and model. The model assumes that memory traces are feature vectors in a representational space that are noisily encoded and maintained, and that retrieval is a sampling process (e.g., Wald, 1947; Edwards, 1965; Shiffrin, 2003) that obtains and integrates noisy evidence of match to a cue. This process updates a posterior distribution over all memory items in parallel until one of the items reaches a (strategically set) threshold (Fig. 1). This mechanism is the Bayes-optimal solution for two hypotheses, and is an approximation to the optimum for multiple hypotheses (Baum et al, 1994). It provides detailed accounts of logarithmic frequency and orthographic neighborhood effects in lexical decision and naming (e.g. Norris, 2006).

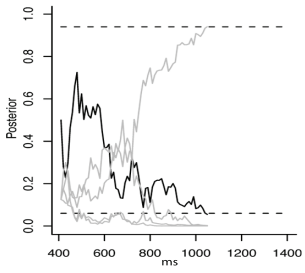


Figure 1. Posterior probabilities in a sampling memory retrieval process.

Similarity-based interference as a signature of optimal discrimination. We show that while there are plausible alternatives to sampling as a cue retrieval mechanism (e.g. race models), these mechanisms have worse speed-accuracy tradeoffs than the Bayesian approximations. This is in spite of the fact that the approximately-optimal mechanisms suffer from similarity-based slowdowns, and the alternatives do not. In other words, interference slowdown is a signature of approximately optimal discrimination in the face of noise, rather than an arbitrary constraint on STM. We ground this explanation mathematically in the normalization step of Bayesian inference.

Simulation results. First, we establish the viability of the model as a general theory of STM by providing single-parameter accounts of RTs and errors in five different STM interference experiments (e.g. Atkins et al 2011). Next, we show that the model recovers key benchmark clausal embedding and interference effects in parsing (e.g. Levy & Gibson, 2013). Finally, we show that the model reproduces patterns of interference in a recent speed-accuracy tradeoff (SAT) study of sentence processing (Fig. 2), and show how the model challenges existing interpretations of those results that assume asymmetries in the effects of syntactic and semantic cues.

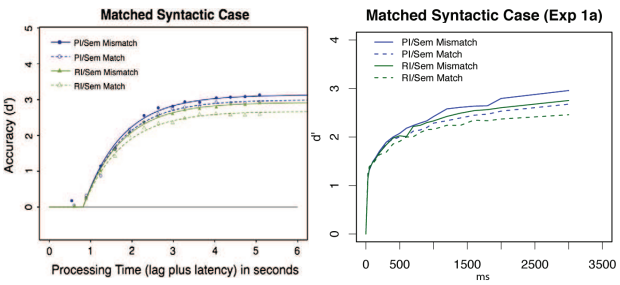


Figure 2. *Left:* Empirical SAT curves as a function of proactive and retroactive semantic interference (Van Dyke & McElree, 2011). *Right:* Model-generated curves under the same experimental conditions.

Integrating perception & memory. To conclude, we sketch a proposal for an adaptive model of reading that integrates saccadic control with a sampling approach to perception and memory. Apart from architectural integration, a key theoretical dividend is that both "forward-looking" surprisal effects (Hale, 2001; Levy, 208) and "backward-looking" interference effects are derived as a consequence of a computationally rational response to memory and perceptual noise.

OV-VO yield imperfect mirror-images: On the impact of length on sentence word order

Idoia Ros, Mikel Santesteban (University of the Basque Country), Kumiko Fukumura (University of Strathclyde) & Itziar Laka (University of the Basque Country)
idoia.ros@gmail.com

What is the impact of constituent length on word order cross-linguistically? Hawkins (2004) claims that minimizing the distance between constituents determines word order preferences to render sentence processing optimally efficient. Specifically, Hawkins proposes the Minimize Domains principle (MiD), which predicts a mirror-image effect for VO and OV languages: VO languages will prefer short-before-long word orders, while OV languages will favour long-before-short orderings. This effect has been found in English (VO) and Japanese and Korean (OV). We tested this hypothesis in two languages with varying degrees of word order freedom: Basque (OV) > Spanish (VO).

24 Basque and 18 Spanish native speakers were presented a series of phrases of varying length in their native language, and asked to arrange them in sentences. Replicating the experimental design in Yamashita and Chang (2001), the length of NPs of transitive sentences (All-Short, Long-DO, Long-S) and of ditransitive sentences (All-Short, Long-DO, Long-IO) was manipulated. According to MiD predictions, Spanish speakers showed a preference to place short constituents before long ones and Basque speakers tended to place long NPs before short ones. However, contrary to MiD predictions: (a) the strength of the preference was stronger in Spanish than in Basque, and (b) the impact of the MiD was stronger in ditransitive sentences (see Table 1).

Table 1. *Percentages of* ***shifted (non-canonical)*** *sentences in Basque (transitive: DO-S, ditransitive: DO-IO) and Spanish (transitive: DO-S; ditransitive: IO-DO) per experimental condition.*

	Length: Transitives			**Length: Ditransitives**		
Language	All-short	Long-S	Long-DO	All-short	Long-IO	Long-DO
Basque	3.6%	5.5%	19.7%	41.8%	33%	60.1%
Spanish	0%	0%	0.8%	10.4%	10.2%	84.9%

In sum, MiD's predictions do not fully hold: Spanish and Basque speakers differ in the frequency and type of sentences where length dependent word order rearrangaments take place. These results are not expected if length is the main factor determining word order, because it would not be expected that the degree of impact that efficient processing has on word order will vary cross-linguistically.

Following MacWhinney (1987), we will argue that when word order is not a reliable processing cue, as in languages with free word order, its efficiency to signal structure is lower. Hence, as freedom of word order increases (Basque > Spanish), the impact of MiD will decrease. More generally, our results show that information given by language-specific cues needs to be taken into account to explain variations in cross-linguistic patterns of sentence word order preferences.

References

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*.
Yamashita, H., & Chang, F. (2001). Cognition, 81, B45–B55.
MacWhinney, B. (1987). *Mechanisms of language acquisition,* pp. 249-308.
Dennison, H. Y. (2008). Working papers in Linguistics: University of Hawaii at Manoa 39(2), 1–16.

A multiple argument overlap boost in Japanese structural priming

Hiroko Yamashita (Rochester Institute of Technology) & Franklin Chang (University of Liverpool)
hxygsl@rit.edu

Structural priming studies have found that overlap in the verb between prime and target can create a lexical boost, where priming is enhanced over abstract syntactic priming (Pickering & Branigan, 1998). These results have been used to argue for a production architecture with verb-structure links. On the other hand, within learning-based accounts of priming (e.g., Chang, Dell, & Bock, 2006), structural priming arises from sequence learning involving lexical and abstract structural regularities. Thus, one prediction of a learning-based account is that priming should also be boosted when there is overlap in non-verb sequences.

To contrast these two accounts, we compared verb-overlap and non-verb (argument) overlap in dative priming in Japanese. Japanese is an SOV language with case-marked arguments, and the dative alternation occurs with scrambling. Because all arguments precede the verb, the effects of arguments are observed more transparently than in SVO languages. Both the canonical IOBJ-OBJ order shown in sentence (1) and the non-canonical OBJ-IOBJ order in (2) convey the meaning "The author explained the summary to the reporter."

(1) Author-TOP reporter-IOBJ summary-OBJ explained.
(2) Author-TOP summary-OBJ reporter-IOBJ explained.

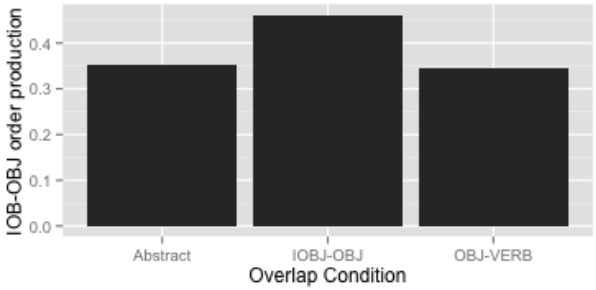
Participants were asked to reproduce a sentence presented by RSVP. Due to the speed of presentation (400 msec/word), the subjects would occasionally flip the order of the two objects. The dependent measure was the proportion of flips that match the structure of the primes.

Primes were always in the IOBJ-OBJ order. Targets were in the OBJ-IOBJ order, as in (2). The overlap between the primes and the target was manipulated by either overlapping the indirect and direct objects with no verb overlap, (3), or overlapping the direct object and the verb with no indirect object overlap (4). These were compared with abstract structural priming, which had no overlap in lexical items, as in (5).

(3) IOBJ-OBJ overlap: Editor-TOP reporter-IOBJ summary -OBJ announced.
(4) OBJ-VERB overlap: Editor-TOP reader-IOBJ summary -OBJ explained.
(5) abstract: Editor-TOP reader-IOBJ novel-OBJ announced.

The results showed no additional priming in the OBJ-VERB overlap condition over priming by abstract condition. Interestingly, a boost in the IOBJ-OBJ condition was found over the abstract condition ($z=2.92$, $p<0.004$).

The lack of a verb-overlap boost in this study is not consistent with architectural accounts of lexical boost. It is not clear, however, that this can be explained as a Japanese-specific effect, since Deng et al. (2013) did find a boost in sentence completion. Another possibility may be that a verb-based lexical boost does not appear in RSVP, as observed in Tooley & Bock (2011). We are conducting an English version of this study and also have found little evidence for a verb-based lexical boost. Crucially, the boost in priming in the IOBJ-OBJ overlap condition suggests that participants retain a memory of the argument sequence that enhances their own structural choices. This supports theories of priming which argue that structural priming results from sequence learning at multiple levels in the language system.



Cross-linguistic differences in processing double-embedded relative clauses: Working-memory constraints or language statistics?

Stefan Frank (Radboud University Nijmegen), Shravan Vasishth (University of Potsdam) & Thijs Trompenaars (Radboud University Nijmegen)
s.frank@let.ru.nl

Double-embedded relative clause structures in English, as in (1), are very hard to comprehend and often judged to be unacceptable. Omitting the second verb results in an ungrammatical sentence that can nevertheless be considered more acceptable, an effect known as the *grammaticality illusion*. Gibson & Thomas (1999) argue that the illusion is caused by working-memory constraints that lead to one of the three verb predictions being forgotten.

(1) The mother who the daughter who the sister found$_{V1}$ frightened$_{V2}$ greeted$_{V3}$ the$_{Det1}$ grandmother$_{N1}$ on$_{Prep}$ the$_{Det2}$ tricycle$_{N2}$.

In a series of self-paced reading (SPR) and eye-tracking experiments, Vasishth et al. (2010) observed slower RTs for grammatical than for ungrammatical versions of English double-embedded structures, while a reversed grammaticality effect (faster RTs on grammatical items) was found with German. They argue that this cross-linguistic difference arises because German speakers are more used to keeping the prediction of an upcoming verb in working memory so that they do not exhibit the grammaticality illusion when a verb turns out to be missing. The same would hold for Dutch, which is also verb-final in relative clauses. We replicated the reversed effect in an SPR study with 24 participants reading 16 double-embedded target sentences (half of which missing a verb) interspersed among 56 fillers (Figure 1, left panel).

If the German/Dutch results are indeed due to an increased ability to retain predicted verbs in working memory, German and Dutch readers should also show the reversed grammaticality effect when reading English. By contrast, the cross-linguistic difference could mirror the difference in frequency of head-final structures. Since these are less frequent in English, sensitivity to the language-specific statistics of word-order patterns will result in a slowdown when three consecutive verbs are encountered in English, but not in German/Dutch. Under this alternative account, German and Dutch readers should show the grammaticality illusion in English, and the illusion should be stronger for readers with more exposure to English.

We tested German (N = 34) and Dutch (N = 29) native speakers on English versions of the Dutch items (a grammatical item is shown in (1)). Both groups displayed the grammaticality illusion: Like English native speakers, they read the word following the main verb more slowly in the grammatical than the ungrammatical condition (Figure 1, center and right). A Bayesian hierarchical linear model showed that this effect was larger for Dutch subjects (Group × Grammaticality interaction: b = 0.04; $P(b > 0)$ = 0.98), who were all students of English and significantly more proficient in English than the German subjects, as indicated by their scores on an English proficiency test. Thus, readers' language-specific sensitivity to word-order patterns, rather than working memory constraints, drives the reversal of the grammaticality effect.

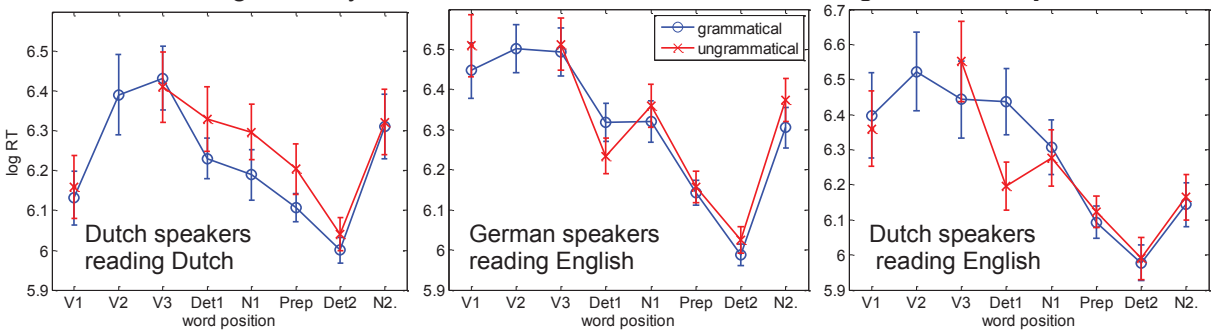


Figure 1: Log-transformed reading times by region on target sentences, from first verb onwards. Error bars indicate 95% confidence intervals.

Effects of Repeated Exposure on Sentence Processing Time: Object Relatives and "Early Closure" Ambiguities

Matthew J. Traxler & Kristen M. Tooley e-mail: mjtraxler@ucdavis.edu

Standard theories of syntactic knowledge treat it as being essentially fixed in adult comprehenders, or as being subject to different learning and memory mechanisms in adults compared to children (Ullman, 2003). However, implicit learning and related approaches to syntactic knowledge view it as more plastic and subject to change (Chang et al., 2003; cf. MacDonald et al., 1994). Recent evidence suggests that exposure changes sentence processing performance over both shorter and longer time scales (Boudewyn et al., in press; Long & Prat, 2008; Traxler & Tooley, 2008; Wells et al., 2009). While we have demonstrations of syntactic priming effects in comprehension, and longer-lasting facilitation of whole-sentence processing, studies to date have not examined long-term effects of exposure on the process of sentence interpretation as it unfolds (but see Tooley, 2009, CUNY).

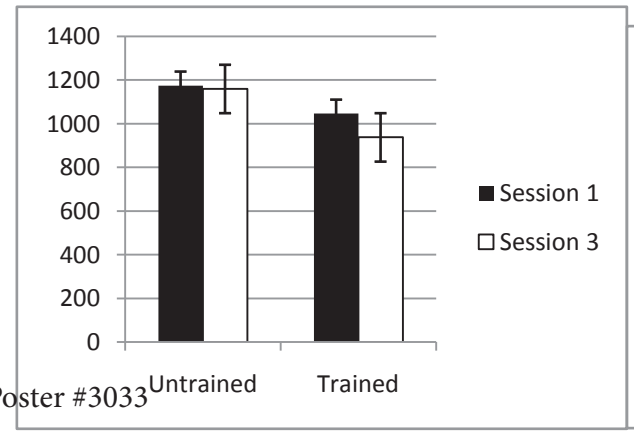
One eye-tracking and one self-paced reading experiment investigated longer-term consequences of exposure to sentences of different types. In both experiments, participants completed three one-hour testing sessions. 24 hours intervened between testing sessions. In the first two sessions, half of the participants received extensive exposure (~80 sentences per session) to one sentence type (the *trained* type), while the other half did not. At the third session, both groups of participants were exposed to sentences of both types. Counterbalancing ensured that a given sentence appeared equally often as a training stimulus and a test stimulus. In experiment 1 (N = 26 participants, 40 items), object relatives like (1) were trained:(1) The attorney *that the banker phoned* filed a lawsuit. (Object Relative)

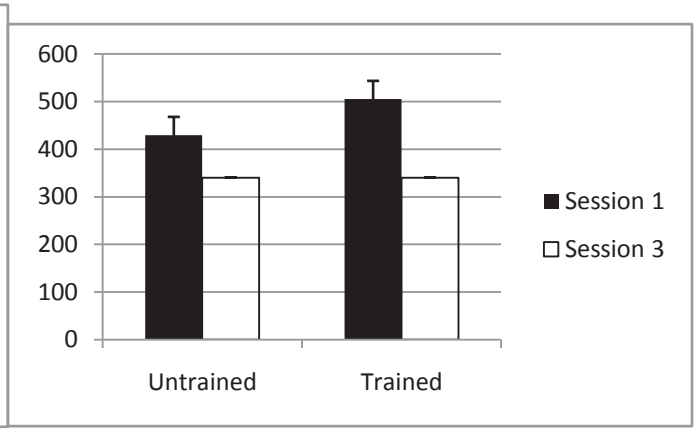
In Experiment 2 (N = 16 participants, 56 items), "early closure" sentences like (2) were trained:(2) *While Fernanda was changing the baby played on the floor*.

In Experiment 1, reduced relatives served as the untrained control. In Experiment 2, "late closure" sentences served as the control.

The eye-tracking (Experiment 1) and self-paced reading data (Experiment 2) were subjected to mixed effects modeling to examine effects of *training* (trained vs. untrained sentence type), *session* (first vs. third session) and their interaction. In Experiment 1, random-intercept models of the relative clause region (*that the banker phoned*) produced interactions of training and session in first-pass and total time [first pass: $t(1203) = 1.94$, $p = .05$, Se = 105.4; total time: $t(1449) = 1.88$, $p = .06$; see left-hand chart]. In Experiment 2, the model indicated an interaction of training and session [$t(735) = 2.11$, $p = .04$, Se = 41.58; see right-hand chart].

Thus, both experiments indicate that exposure to sentences of a given type can facilitate processing of new sentences of that type. The findings from Experiment 1 confirm prior self-paced reading results (Wells et al., 2009). The current results extend prior results by showing that facilitation takes place as syntactically challenging portions of a sentence are being processed. Experiment 2 indicates that exposure effects extend to sentences containing an "early closure" ambiguity.





Structural Forgetting in Processing Complex Mandarin Relative Clauses

Yuxia Wang (Shanghai Jiaotong Univ.), Francois Rigalleau & Manuel Gimenes (Univ. of Poitiers)
yuxiawang@sjtu.edu.cn

The processing of complex relative clauses has been a focus in the resource-based theories of sentence comprehension. When the human parser is dealing with complex sentences, it might reach the verge of extreme working memory load and selectively forget certain element and treat a structural ungrammatical sentence as grammatical. The structural forgetting effect has been reported with European languages (e.g., French, Gimenes, et al. 2009; English, Gibson & Thomas, 1999), where they found that the grammaticality illusion stays with the most embedded verb in the double-embedded relative clause (RC).

This study chooses a typologically different language, Mandarin Chinese, as the target language. Mandarin is a SVO language with a head-final RC structure. This presents two major differences from English and French: 1) the nested dependency relation exists with the subject RC type, and 2) Mandarin requires the retention of prediction on NP in processing a relative clause, (i.e., VO $DE_{relative\ marker}$ N). The purpose is to investigate whether the grammaticality illusion driven by VP forgetting in European languages could also be observed in NP missing in Mandarin. Given that locality-dependency hypothesis makes no difference concerning different lexical categories, we hypothesize that the NP missing effects could be founding Mandarin. To achieve this, the study conducts two offline complexity judgment tasks and one online self-paced reading task (SPRT). The general results support our hypothesis.

Using a 2x2 design, study 1 compared RC types (subject RC versus object RC) and sentence completion status (with or without NP3) in an offline complexity-rating task. An interactive effect was reported, F1(1,39)=20.26, p<.0001; F2(1,31)=10.52, p<.01. Further analysis showed that native Mandarin speaker rated the NP3_missing condition as more difficult to comprehend only when it occurred in object RCs, not in subject RCs. To have a more stringent comparison of the NP omission effect in subject RCs, study 2 compared 3 conditions: All_NPs, NP3_missing and NP4_missing. ANOVAs revealed that the missing NP3 was rated significantly easier than the All_NPs condition (F1(1,29) =7.08, p<.05; F2(1,23)=5.26, p<.05) and the missing NP4 (F1(1,29)=18.36, p<.0001; F2(1,23) =15.2, p<.001). Meanwhile, the NP4_missing was judged significantly more difficult than the All_NPs condition only by subjects (F1(1,29)=4.41, p<.05), but not by items. The results suggest that the effect of structural forgetting is clearly observed in the missing NP3. Such an effect is not found in the missing NP4 condition, where participants judges it as more difficulty, maybe because of the difficulty to integrate the last NP in the matrix clause level. Study 3 was a moving-window SPRT study and focused on three conditions: All_NPs, NP3_missing and a grammatical, control condition of NP3_missing. The results of the online complexity rating task confirmed the above results; however, the reading time data at the critical region suggested that NP3_missing condition took significantly more time than the All_NPs (F1 (1,35)=5.70, p<.05; F2(1,32)=7.70, p<.01) and the control condition (F1(1,35)=7.02, p<.05; F2(1,32)=7.65, p<.01). This suggests that Mandarin speakers maintain the online NP3 prediction, consistent with Vasishth's (Vasishth, et al. 2010) observations about on-line processing of VP2 omission in German. The current work provides new evidence in corroboration of the Syntactic Prediction Locality Theory (Gibson, 1998) concerning off-line judgments. It also shows that these judgments cannot be taken as proofs for an insensitivity of the parser to omissions.

References: 1. Gibson, E. (1998). *Cognition, 68,* 1-76. 2. Gibson, E., & Thomas, J. (1999). *LCP, 14,* 225-248. 3. Gimenes, M. et al. (2009). *LCP, 24(3),* 440-449. 4. Vasishth, S. et al. (2010). *LCP*. 25(4):533-567.

Prototypical Stimulus：All_NPs: NP1 VP1 VP2 VP3 NP2 NP3 NP4

[yiyuan pingqingle [e_i zhiyu [[e_j shuaiduan yitiaotui de] bingren$_j$] de] yisheng$_i$].
Hospital hired cured broke one-leg REL patient REL doctor.
The hospital hired the doctor who cured the patient who broke one leg.

NP3_missing: [yiyuan pingqingle [e_i zhiyu [[e_j shuaiduan yitiaotui de] ~~bingren$_j$] de]~~ yisheng$_i$].
NP4_missing: [yiyuan pingqingle [e_i zhiyu [[e_j shuaiduan yitiaotui de] bingren$_j$] ~~de] yisheng$_i$~~].

Anti-locality preference in the processing of Japanese reflexive binding

Akira Omaki (Johns Hopkins University), Brian Dillon (UMass Amherst), Takuya Kubo, Manami Sato, & Hiromu Sakai (Hiroshima University
omaki@jhu.edu

An emerging generalization about the online comprehension of direct object reflexive anaphors (e.g. English *himself* or Mandarin Chinese *ziji*) is that comprehenders have rapid and reliable access to the local subject for purposes of reflexive binding. One source of evidence for this is the lack of interference effects from syntactically inaccessible antecedents in languages with strictly local reflexives, like English [1]. A second source of evidence comes from languages with long-distance (LD) reflexives, which can be grammatically bound by local or non-local subjects. When processing LD reflexives in Mandarin, comprehenders prioritize the local subject as an antecedent [2,3]. This online locality effect is widely considered to follow from architectural constraints on memory retrieval [4]. The present eye-tracking study explores whether the rapid access to the local subject may be tied to VO word order in languages like Mandarin. As argued by [5], processing a verb may restore the local subject to focal attention. If so, the post-verbal position of the reflexive in Mandarin may cause the locality preference observed by [2,3]. We test this possibility by examining the LD reflexive *zibun* in Japanese, a verb-final OV language. We report evidence for an *anti-locality bias* in Japanese reflexive binding, in contrast to Mandarin. We argue that the order of verb and reflexive can have a large impact on binding mechanisms, and that the post-verbal position of the reflexive may be critical in driving the online locality effect seen for Mandarin LD reflexives.

As *zibun* requires an animate and singular antecedent, the eye-tracking experiment (n=32) used 16 sentence sets (1) that manipulated the animacy or number feature of the local and non-local subject. If *zibun* is susceptible to a locality bias (as in Mandarin), reading disruption is expected at *zibun* in the local mismatch, LD match condition (2a), compared to the local match, LD mismatch condition (2b). On the other hand, an anti-locality bias predicts an opposite reading time pattern. Offline, norming survey data confirmed that Japanese speakers indeed preferred the animate antecedent for *zibun* (>86.5% on average). The target sentences included a) a modifier clause (e.g. “before the negotiation failure”) between the two subjects in order to reduce the cost associated with having two adjacent subjects in verb-final languages [6], as well as b) an adverb (e.g. “somehow”) between N2 and the critical reflexive to ascertain that the local subject does not remain in focal attention at the critical anaphor.

No reliable main effect was observed in first fixation or first-pass durations in the critical region. The go-past duration revealed a significantly longer fixation at the post-reflexive, spill-over region in (2b) compared to (2a), consistent with an anti-locality preference for *zibun*. In accordance with the LD mismatch cost, the total fixation duration on the N1 region was significantly longer in (2b) compared to (2a). This is likely to reflect re-reading of the N1 region in LD mismatch condition (2b) given the lack of main effects in early measures in this region.

Taken together, this study demonstrates a novel finding: the Japanese LD reflexive *zibun* is preferentially associated with the non-local subject in comprehension, and thus is an example of an online anti-locality preference for reflexive. This suggests that the word order between the verb and the reflexive is a critical factor that supports the rapid and reliable access to the local subject in head-initial languages.

(1)*[N1]-ga, koushoushippai-no maeni, [N2]-ga nazeka* ***zibun-o*** *kashinshiteita-to tsutaeiteita.*
N1-Nom negotiation failure-Gen before N2-Nom somehow self-Acc overtrusted-that reported
“[N1] reported that before the negotiation failure, [N2] was somehow over-confident in himself.”

	N1 (LD, non-local subject)	N2 (local subject)
(2a)	Match: *kokuren-no choukan ”UN diplomat”*	Mismatch: *ooku-no kiji* ”many articles”
(2b)	Mismatch: *ooku-no kiji* ”many articles”	Match: *kokuren-no choukan ”UN diplomat”*

[1] Sturt 2003, *JML, 48*. [2] Chen et al., 2013. *Mind Research Repository.* [3] Dillon et al. 2010, CUNY Talk. [4] Lewis et al., 2006. *TICS*, 10. [5] King et al, 2010. CUNY Poster. [6] Uehara 1999. PhD Thesis, CUNY Graduate Center.

Local coherence effects and crossing reflexive- and wh-dependencies

Michael Frazier, Peter Baumann, Lauren Ackerman, David Potter & Masaya Yoshida
(Northwestern University)
fraze@u.northwestern.edu

Background:

In real-time sentence processing, the human parser is sometimes susceptible to 'local coherence' effects. In these situations, the parser may disregard the syntactic requirements of the larger context and produce parses of 'locally coherent' portions of the input string that cannot be integrated into the whole sentence [Tabor et al. 07; Konieczny et al. 09].

This study aims to determine whether the parser attempts to establish locally-coherent, but globally impossible, reflexive dependencies when they are interrupted by the tail of a long-distance wh-dependency. Interrogated embedded infinitive clauses (1a) are an ideal testing ground for the behavior of crossing wh- and reflexive dependencies because, for some classes of infinitive-complement taking verbs, omitting a sentence-initial wh-word yields a grammatical sentence with a different interpretation (1b). Thus examples like (1a) contain a local coherence from 'did' to the sentence end.

1a. Which man did John expect to have injured himself?

1b. Did John expect to have injured himself?

A reflexive must normally corefer with its closest potential antecedent: *himself*=*John* in (2a). The pattern in (2b) indicates that when a wh-phrase is moved from after expect, it must be the antecedent of himself instead (*himself*=*which man*). The results of an acceptability judgment experiment support this: the example is acceptable when *which man*, but not *John*, is the antecedent of *himself* (n=80;p<.05).

2a. The man expected John to have injured himself.

2b. Which man did John expect to have injured himself?

If the wh-dependency is resolved in the middle of the potential local coherence (at approximately the position of expect), then the tail of the wh-dependency may interrupt a potential local coherence, like a normal NP intervening between the subject of the locally coherent parse (*John* in 2b) and the reflexive, so himself may take the wh-word as its antecedent.

Experiment (n=24):

An eye-tracking text reading experiment used the paradigm (3). In (3), gender congruency (factor 1) of the grammatically acceptable antecedent (the wh-phrase) and (factor 2) of the intervening, locally coherent but globally unacceptable antecedent (the subject NP) with the reflexive were manipulated as independent factors in a 2x2 factorial design. Using the paradigm in (3), we investigated whether wh-dependency resolution interrupts the potential local coherence. If so, the wh-phrase should be chosen on-line as the antecedent of the reflexive in (3). If the locally coherent parse is chosen, the closest overt NP John should be initially interpreted as the antecedent. Which antecedent the parser chooses can be measured via gender-mismatch effects [e.g. Sturt 2003].

(3) Which cowgirl did Anna/Steven expect to have injured himself/herself due to negligence?

Analysis of regression path time reveals significant results. There was a main effect of the accessible (wh-phrase) antecedent's gender congruency with the reflexive (ps<.05). That is, the reflexive region was read significantly slower in Wh-phrase Mismatch condition than in the Wh-phrase Match condition. On the other hand the subject NP (Anna/Steven in 2) did not affect the reading of the reflexive (regression path; ps<1).

Conclusion:

This result indicates that the parser selects the linearly further but globally grammatical wh-phrase as the antecedent of the reflexive in online reading, rather than the intervening subject NP which is a potential locally coherent antecedent for reflexive binding. This result strongly suggests that the resolution of a wh-dependency suffices to interrupt local coherence and permit the parser to build the globally grammatical structure immediately.

Locality and anti-locality effects in the processing of expected and unexpected inputs: A study of NPI dependencies in Japanese

Kentaro Nakatani (Konan University)
kentaron@konan-u.ac.jp

The present study deals with the question of how "dual" grammatical dependencies such as dependencies between an NPI (Negative Polarity Item) and Neg are tracked in the on-line processing of Japanese, a head-final language. These dependencies are "dual" in that higher-order grammatical relations (e.g., NPI=Neg, *wh*=Q) must be tracked along with the tracking of base-line theta-relations (e.g., NPI=V, *wh*=V). The main concern here is the locality effects (Gibson, 1998, 2000; Levy & Keller, 2013; etc). Suppose that exactly the same words precede the V, and the position of an NPI relative to the V is varied by scrambling, as in (1) below (*-sika* is an NPI marker meaning "anyone/thing but"; *-ga* is a nominative marker; CC represents a complement clause):

(1) a. SIKA-DISTANT: NP-*sika* [CC] V-NEG c. GA-DISTANT: NP-*ga* [CC] V-NEG
b. SIKA-LOCAL: [CC] NP-*sika* V-NEG d. GA-LOCAL: [CC] NP-*ga* V-NEG

An activation decay model (Vasishth & Lewis, 2006) may predict locality effects for the SIKA conditions, because it might be the case that nothing activates *-sika* until the NEG is encountered. Nakatani (2009) reports a finding supporting this prediction.

The present study modifies Nakatani's design (GA/SIKA x DIS/LOC) by adding another factor, the NEGative/AFFirmative factor. For NP-*ga*, an encounter with negated V is marked; for NP-*sika*, an input of affirmative V is highly marked—i.e., ungrammatical. The question here is how such "marked encounters" are processed, especially in relation to the locality factor. One possibility is that they are processed just like unmarked encounters (with a slightly heavier load), showing locality effects. Another (rather novel) possibility is that the marked encounters may show anti-locality effects—based on the assumption that the memory component not only stores retrieval targets, but also predicted heads (à la Gibson's 1998 Syntactic Prediction Locality Theory), both of which are prone to decay (cf. Gibson & Thomas, 1999). It thus can be predicted that the marked encounters (NEG for GA, and AFF for SIKA) may give rise to anti-locality effects, in proportion to the degree of the decay of stored predictions.

A moving-window, self-paced reading experiment (32 items/74 fillers) was conducted with 67 native speakers of Japanese to test these predictions. For the NEG conditions (Fig.1), the RTs at the critical region (=negated V) revealed a significant interaction between the two factors ($p < .01$), and for the AFF conditions (Fig.2), again, an interaction was observed ($p < .01$) at the critical region (=V), with pairwise comparisons within the GA conditions and within the SIKA conditions both revealing significant contrasts between the DISTANT and LOCAL conditions ($p < .005$; $p < .05$; respectively). Crucially, anti-locality effects were found for marked, unexpected heads, while locality effects were observed for expected inputs. This supports the "prediction storage" component in the working memory, related to but distinct from the probabilistic expectation component proposed by Hale (2001) and Levy (2008).

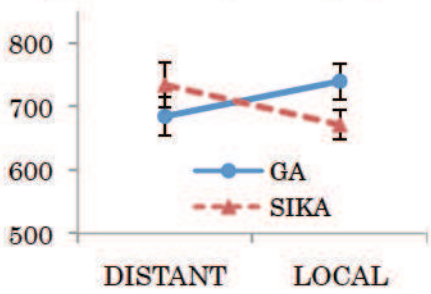


Fig.1 Mean RTs at V in the NEG conditions

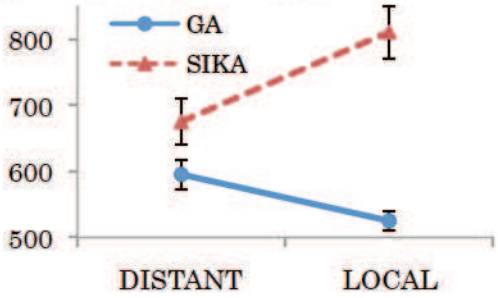


Fig.2 Mean RTs at V in the AFF conditions

Effects of the second language on syntactic processing in the first language

Giuli Dussias, Lauren Perrotti, Matt Brown (Penn State University) & Luis Morales (University of Granada, Spain)
pdussias@psu.edu

Past research has shown that adult native speakers of Spanish immersed in an English-speaking environment adopt processing routines of their second language (L2) when processing their first language (L1). Here we ask whether changes in processing routines can be triggered by overexposing bilingual speakers to particular structures so that bilinguals who have undergone changes when processing in their L1 'move back.' We hypothesized that if the parser's configuration is related to language exposure (e.g., MacDonald & Seidenberg, 2006), bilinguals' parsing preferences are expected to change as a function of the frequency with which the relevant structure appears in an experimental session. We investigated this in the context of temporarily ambiguous relative clauses as in *Arrestaron a la hermana del hombre que estaba enferma* (Someone arrested the sister$_{FEM}$ of the man who was ill$_{FEM}$). Here, the relative clause *que estaba enferma* (who was ill) can attach to the higher noun (*hermana/sister*) or the lower noun (*hombre/man*) in the complex noun phrase. Because the adjective *enferma* (ill) is marked with feminine grammatical gender, the correct interpretation is one where the relative clause attaches to *hermana* (also marked with feminine gender). Spanish-English bilingual speakers living in an English-speaking (US bilinguals) or a Spanish-speaking (Spain bilinguals) environment were recruited. The study involved three phases. In phase 1, we carried out an eye-tracking study to determine the participants' initial attachment preferences in their L1 (Spanish for both groups of speakers). In Phase 2, they participated in an 'intervention' study that exposed them to a biased sample of 120 relative clause constructions in their second language (English) over a 5-day period. Participants who showed initial attachment preferences in Spanish that favored low attachment received a biased sample of high attachment sentences; those whose initial attachment preferences favored high attachment received a low attachment treatment. In phase 3, eye-movement records were again collected to determine participants' attachment preferences after the intervention study. The Spain bilinguals were tested two days after the intervention study had ended. The US bilinguals were, in addition, tested a week later to assess non-immediate effects of exposure. At pre-test, all participants were also administered the AX-CPT task (to measure sustained attention) as well as the Raven's matrices and a standardized English language test (to match the US and Spain bilinguals). Preliminary results reveal the following: (1) Changes in attachment preference occurred in both groups of speakers but were modulated by performance in the AX-CPT task; (2) "High attachers" showed evidence of low attachment preferences at post-test 1 and at post-test 2 despite the fact that the input in the intervention study was in their L2, and even when they were immersed in an entirely Spanish-speaking environment. This provides evidence for a high level of permeability between the bilinguals' two linguistic systems; and (3) "Low attachers" showed evidence of a high attachment preference only at post-test 2, suggesting that learning to attach high requires consolidation of information. The findings will be discussed in terms of the role of statistical learning in sentence comprehension processes (e.g., Gennari & MacDonald, 2009; Wells et al., 2009) and the implications of the results for theories of cognitive control.

References

Gennari, S. P., and MacDonald, M. C. (2009), Linking comprehension to production: The case of relative clause. *Cognition*, 111, 1-23.

MacDonald, M. C., Seidenberg, M. S. (2006). Constraint satisfaction accounts of lexical and sentence comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), Handbook of Psycholinguistics, 2nd Edition (pp. 581-611). London: Elsevier, Inc.

Wells, J., Christiansen, M., Race, D., Acheson, D., MacDonald, M. C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*, 250-271.

A chunk of coffee: an event-related brain potential (ERP) study on the processing of Mandarin classifiers

Zhiying Qian & Susan M. Garnsey (University of Illinois at Urbana-Champaign)
zqian3@illinois.edu

Classifiers (CLs) are words that categorize nouns based on their features (e.g., Mandarin *tiao*- CL for long, soft and bendable things; *tou-CL* for big animals). In English, classifiers are nouns that are used primarily with mass nouns (e.g., “cup” in “a cup of coffee”, “sheet” in “a sheet of paper”). In Mandarin, in contrast, classifiers must be present whenever any noun is quantified (e.g., three *zhang-CL* desk) or specified (e.g., this/that *zhang-CL* desk). Mandarin classifiers are more like function words that combine with quantifiers/specifiers rather than content words that can stand alone, and they must be congruent with the meaning of the noun that follows them (e.g., it does not make sense to say “three *tai-CL* chair” as “tai-CL” means “machine-like”, but “three *tai-CL* computer” is good).

Some linguists have argued that classifiers should be considered syntactic entities, based on the observation that some classifiers are completely redundant with the meaning of the noun, (e.g., Greenburg, 1972), but others (e.g., Aikhenvald, 2000) have argued that classifiers are semantic units that set selectional restrictions for the set of nouns they can be used with. One possible source of evidence about the nature of classifier processing is event-related brain potential (ERP) responses to classifier constructions, since different ERP components have been found to respond to semantic and syntactic aspect of language processing. Two previous studies on Japanese classifiers that manipulated the congruency between a classifier and the noun following it have reported conflicting results: Mueller et al. (2005) found a Left Anterior Negativity (LAN) in response to mismatching classifier-noun pairs (e.g., “cat” in “a bird-like-CL cat”), suggesting increased difficulty in morpho-syntactic processing, while Sakai et al. (2006) instead found an N400 effect, suggesting greater difficulty in semantic processing. The present study is intended to further elucidate the nature of the processing of classifier-noun combinations.

Sentence stimuli included versions with classifier-noun matches and mismatches. In addition, the degree of specificity of the classifier was manipulated. Some classifiers are much more restrictive than others, allowing only a limited number of nouns to follow them (e.g., “zhan-CL” can only take nouns that are lamps) while other classifiers can take many types of nouns (e.g,. “tiao-CL” can go with towel, river, road, leg, news, rope and many more). Classifier specificity were normed. Native Mandarin- speaking participants (n=33) read Mandarin sentences one phrase at a time at a rate of 400 msec per phrase. All experimental items followed the pattern Location Phrase, Number+Classifier, Noun, Sentence Completion, as in: *On the table a cup/chunk of coffee is already cold.* Number and classifier were presented together as one region and the critical noun was presented separately immediately after it.

Responses to the number+classifier region showed a frontally distributed P200 (100-300 msec) specificity effect, with larger P2 in response to specific than to general classifiers, consistent with other findings that P200 amplitude increases when context is highly constraining (Wlotko & Federmeier, 2007; Lee, Liu, & Tsai, 2012). The responses to the nouns following the classifiers were characterized by a centro-parietally distributed N400 (300-500 msec) congruency effect, with nouns that mismatched their classifiers evoking a larger N400 than matching nouns. There was no interaction between correctness and specificity in any time-window. These results suggest that the nature of classifier processing in Mandarin is primarily semantic, consistent with Sakai et al.’s (2006) results for Japanese. They also add to a growing body of evidence that the amplitude of the P200 ERP component is sensitive to the amount of constraint provided by linguistic context.

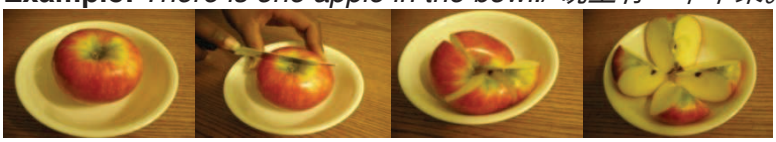
Attuning to cohesion: English count-syntax, the Mandarin general classifier *ge*, and wholeness

Jessica Harding & Chien-Jer Charles Lin (Indiana University)
jeshardi@indiana.edu

I. Introduction. A significant body of research suggests that adults from count-syntax and numeral classifier language backgrounds often attune to different features when asked to interpret the salient qualities of physical entities (Imai & Gentner, 1997; Imai & Mazuka, 2007, Kuo & Sera, 2009, Lucy & Gaskins, 2001). The purpose of the present study is to examine how speakers from both language backgrounds will interpret items with respect to cohesion, or 'wholeness'. Some have claimed that the Mandarin general classifier GE is analogous to English count syntax in that it is semantically 'empty', and merely acts as a syntactic place holder in numeral NPs (Aikenvald, 2000; Beckwith, 2007; Cheng & Sybesma, 1998; Erbaugh, 2002; Myers, 1999). If GE is devoid of semantic content, then the phrase *'yi ge x'* should fall under the same constraints for cohesion as the English count-syntax phrase 'one x'. If *'yi ge x'* has different constraints than 'one x', it is evidence that the two systems have different constraints with respect to cohesion, or 'wholeness' (all parts of an object being contiguous).

II Method. To ascertain how English and Chinese syntactic individuation lead language users to respond to cohesion violations differently, participants in this study completed a picture description judgment task. Participants (16 native speakers of English and 16 native speakers of Mandarin Chinese) rated the appropriateness (from 1, not appropriate, to 5, very appropriate) of the descriptive sentence "*There is one [item] on the plate/ Panzi shang you yige [item] (plate-on have one classifier [item])*" for four photos in different stages of cohesion violation.

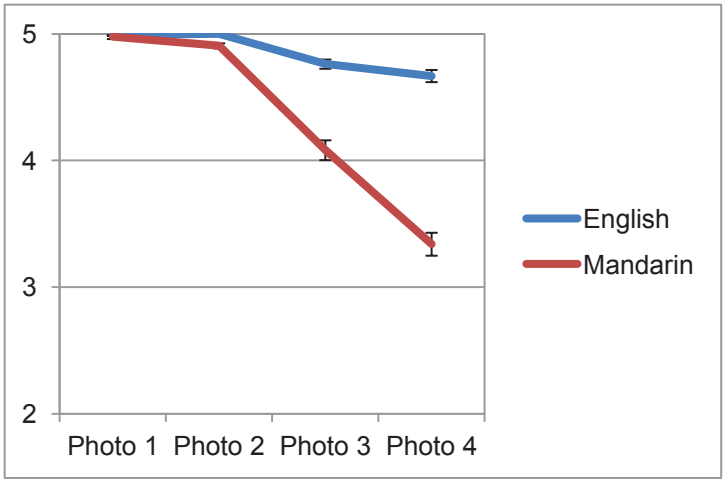
Example: *There is one apple in the bowl./ 碗里有一个苹果。*



Conditions: Photo1-intact, Photo2-immenent dissection, Photo3-close dissection, Photo4-distant dissection

Target items: plum, apple, peach, pear, watermelon, sandwich, doughnut, cake, orange, grapefruit, pie, cantaloupe, kiwi, tomato, potato, pumpkin

III. Results and Discussion. Our results show that there was a significant effect of language, with English speakers being more likely to accept a count-syntax description when foods are not in a cohesive state than Mandarin speakers [$F(1, 30) = 14.4$, $p < .01$], a significant main effect of cohesion violation across different photos [$F(3, 1995) = 253.16$, $p < .001$], and a significant interaction between main effects [$F(3, 1995) = 106.99$, $p < .001$]. Mandarin speakers rated photos 3 and 4 as significantly less appropriate for the description than English speakers ($ps < .01$). These findings suggest that Mandarin classifiers, as an individuating device, encode more salient information about 'wholeness' than count syntax in English. These results have implications for the different cognitive constraints associated with classifiers and mass/count syntax.



Implicit prosody primes appropriately intonated auditory probes

Anouschka Foltz (Bielefeld University), Cody Linne & Shari R. Speer (Ohio State University)
anouschka.foltz@uni-bielefeld.de

Implicit prosody, a prosodic contour projected onto sentences during silent reading, has been evoked to explain various effects in silent reading studies. Implicit prosody effects are associated with an auditory image or "inner speech" in the mind of the reader. However, there is no direct evidence tying a *sentence-level* auditory image directly (i.e. within the same trial) to the measured effects. In addition, little is known about how implicit prosody '*sounds'* - that is, what abstract phonological components of intonation it may contain. We addressed these issues in a visual-to-audio cross-modal priming experiment.

Twenty-seven native English speakers participated. In each of the 32 target trials, participants first saw a pair of sentences involving a corrective contrast either in subject or verb position. (The 32 filler trials involved contrasts in other sentence positions.) Once they had silently read and understood the sentences, they heard an auditory probe word produced either with a salient contrastive L+H* accent (BELINDA) or with no pitch accent (BELINDA) and had to decide as quickly as possible if the probe word was the first word in the second sentence (the target word, e.g. *Belinda*, see (1)). After the experiment, participants read aloud an unrelated text that contained opportunities to produce contrastive prosody.

(1) The four experimental conditions (2x2 design)

	Subject Contrast	**Verb Contrast**
Visual sentence pair	*Jacquelyn didn't pass the test. Belinda passed the test*	*Belinda didn't fail the test. Belinda passed the test.*
Auditory Probe *match*	L+H* (BELINDA)	no pitch accent (BELINDA)
mismatch	no pitch accent (BELINDA)	L+H* (BELINDA)

To account for individual differences in implicit prosody, participants were first grouped based on general characteristics of their overt prosody, i.e. based on whether they produced L+H* followed by no accent vs. H* !H* in potential contrast sentence locations of the unrelated text. Twelve participants each made up the L+H* and H* user groups. (Three further participants were excluded.) We predicted that the L+H* user group would respond to the manipulation of the cross-modal priming task: If these participants generated an implicit L+H* accent on the target word in the *subject contrast* condition, they would respond faster to L+H* probe words (e.g. BELINDA, see (1)) than to no accent probe words (e.g. BELINDA). If they also generated no implicit accent on the target word in the *verb contrast* condition, they would respond faster to no accent probe words (e.g. BELINDA) than to L+H* probe words (e.g. BELINDA).

We fit mixed-effects models with response time as response variable, the predicted factors as fixed effects, and subjects and items as random effects. Redundant fixed effects were removed from the model. Random slopes were added if they improved model fit. An initial analysis revealed a reliable 3-way sentence contrast x probe prosody x participant group interaction. We therefore conducted separate analyses for each participant group with sentence contrast, probe prosody, and the interaction as fixed effects: As expected, only the L+H* user group showed a reliable interaction (*estimate* = 24.800, $t = 2.985$, $p < .01$). Additional analyses showed that they responded faster to '*matching*' compared to '*mismatching*' auditory probes. In addition, both groups responded faster to probes following verb contrast compared to subject contrast sentence pairs (*estimate* = -28.218, $t = -2.116$, $p = .059$ for L+H* users and *estimate* = -25.16, $t = -2.416$, $p < .05$ for H* users). The results present direct evidence that some readers (i.e. the L+H* users) generated implicit L+H* accents on corrective contrast items during silent reading. We thus find initial direct evidence for a sentence-level implicit prosodic contour that includes L+H* accents in critical sentence locations. The results also suggest that there are individual differences in implicit prosody that future studies should address.

Prosodic and syntactic effects in Gapping interpretation

Michael Frazier (Northwestern University) and Katy Carlson (Morehead State University)
fraze@u.northwestern.edu

Gapping is a topic of controversy in the theoretical syntax literature. Researchers differ on the availability of Gapping that crosses a clause boundary, as in (1). Examples of this kind ("long-distance gapping") have been reported as fully acceptable (Pesetsky 1982, Richards 1997) and as fully unacceptable (Jayaseelan 1990, Abe & Hoshi 1995). Furthermore, some approaches predict that complementizers should not occur in the antecedent of the missing portion of a gapping sentence (e.g. Wyngærd 1993, Johnson 1996/2003). We report on a forced-choice interpretation study investigating whether these readings are accessible and grammatically constrained, and find effects of accent placement and complementizer presence.

Materials consisted of 16 ambiguous gapping sentences like (1a), using finite embedded clauses because this is the type of Gapping most widely considered impossible (e.g. Neijt 1979). Examples like (1a) may have two possible readings, the preferred *downstairs* Gapping reading (1b) and the controversial *upstairs* reading (1c). Contrastive accent was placed on either the first subject (*general*) or the second subject (*janitors*), to bias comprehenders to either the upstairs reading or the downstairs reading. This manipulation was crossed with presence vs. absence of complementizer 'that' for a 2X2 design. If upstairs Gapping readings are not available, accent placement should not matter, and upstairs responses should be negligible. If they are available, however, and if focal stress influences Gapping interpretation, participants should arrive at the upstairs reading more often with the first subject accented. If the complementizer 'that' cannot occur in the missing portion of upstairs Gapping, there should be an interaction, with accent on the first subject creating more upstairs responses only without 'that.'

In an auditory-presentation forced-choice comprehension paradigm, 48 participants listened to ambiguous Gapping sentences like (1). On each trial, a participant was then visually presented with one of two yes/no comprehension questions: either one that is true on the upstairs reading and false on the downstairs reading, or vice versa. Participants' responses were analyzed by mixed-effects logistic regression, controlling for which question type they received.

There was a significant overall bias towards the downstairs interpretation across all conditions, with no more than 25% upstairs responses in any condition. This is consistent with earlier work (Carlson et al. 2005) indicating a cost for larger stretches of gapped material. Accenting the second subject yielded more downstairs interpretations than accenting the first subject ($p<0.001$). Conditions with 'that' yielded fewer downstairs interpretations than those without ($p<0.001$). There was also a significant interaction between these effects ($p<0.05$). Post-hoc pairwise comparisons revealed that these effects were driven by the differences between the two non-'that' conditions ($p<0.001$) and between the 'that' and non-'that' conditions with stress on the second subject ($p<0.05$). The presence of ‘that’ reduced the rate of downstairs interpretations vs. conditions with stress on the upstairs subject. That is, upstairs readings were less dispreferred in the presence of 'that', contrary to the predictions of any theory in which gapping cannot delete a complementizer. An explanation of this pattern in terms of a theory of Gapping involving both movement and deletion is discussed. These results strongly suggest that (A) long-distance gapping readings are available, though dispreferred; (B) comprehenders are sensitive to accent placement when parsing an ambiguous gapping sentence; and (C) long-distance gapping is compatible with a complementizer in the missing portion of the gapped clause.

1a. The general supposes (that) the janitors swim at the gym and the writer at the beach…
1b. … and [the writer ~~swims~~ at the beach]…
1c. … and [the writer ~~supposes the janitors swim~~ at the beach]…

When phonological systems clash: L1 phonotactics vs. L2 assimilation

David Li & Elsi Kaiser (University of Southern California)
lidc@usc.edu

Phonological variation may lead to lexical ambiguities. For example, in sentences like 'A quick run picks you up' (Gaskell & Marslen-Wilson, 2001), 'run' may assimilate to 'rum', especially in casual/fast speech (cf. Magnuson et al., 2003, McQueen et al., 2009). Prior work shows listeners compensate for coarticulation; identical acoustical signals in different phonetic contexts can be perceived differently (e.g., Mann, 1980). However, the question of whether adult second-language (L2) learners (L2'ers) compensate for coarticulation like native speakers (L1'ers) has received less attention. When adults acquire a new (non-native) phonological system, how do the two systems compete/interfere? We tested how L1 Mandarin speakers learning English compensate for assimilation, compared to L1 English speakers. Mandarin allows us to study the effects of L1 phonotactics: Since Mandarin phonotactics does not allow words to end in the nasal coda [m], we examined the extent to which the L1 (Mandarin) system influences whether/how listeners compensate for potential assimilation of [n] to [m] in the L2 (English).

PERCEPTION STUDY: We conducted a visual-world eye-tracking study in English (16 targets, 38 fillers). Target sequences (noun+verb) were spliced into carrier phrases. The **speech rate** of *carrier phrases* was manipulated (fast/slow), to change the likelihood of coarticulation. Crucially, *the target sequences were **identical** in the different conditions,* with the sentences differing only in speech rates of the *carrier phrases*. (Target sequences were elicited at a moderate speech rate and contained potentially ambiguous [-m/-n] sounds; confirmed by L1 English responses, below). Participants (16 L1 English; 16 L1 Mandarin/L2 English with high L2 proficiency) were taught the nonwords were names of aliens. They listened to sentences and saw displays with aliens, including one alien whose name is the unassimilated form (e.g., Vone) and another alien whose name is the assimilated form (e.g., Vome). Critical alien names occurred before verbs starting with bilabial stops (a context where assimilation could occur). The task was to mouse-click on the alien mentioned in the sentence (i.e., the alien with the unassimilated name or with the assimilated name, depending on listeners' interpretation).

PREDICTIONS: Based on our earlier work, we expect native speakers to consider the unassimilated form (e.g., Vone) more in the fast-rate than slow-rate contexts (recall that target sequences are constant, only the rate of carrier phrases is manipulated). Will L2'ers show the same pattern, or will the phonotactics of their L1 system 'short-circuit' this process?

RESULTS: L1 and L2 speakers' ***offline** mouse-click choices* differ ($p<.05$), with L2 speakers showing a strong preference for unassimilated forms (but no rate effect for L1 or L2 speakers). This suggests that the Mandarin L1 phonological system weakens L2'ers' consideration of assimilated forms in English, presumably due to the violation of Mandarin phonotactics. **Eye-movement patterns show** that compensation for coarticulation by L1'ers is influenced by contextual speech rate: L1 speakers looked at unassimilated forms more in fast-rate than slow-rate contexts ($p's<.05$ in 600-900ms time-window after noun onset). Importantly, L2 speakers show the same behavior, but later: they also looked at unassimilated forms more in the fast-rate condition, though at a later time ($p's<.05$, 1100-1200ms after name-onset).

DISCUSSION: Our findings highlight the benefits of using online measures such as eye-movements in both L1 and L2 research. Offline mouse-clicks show that L1 phonotactics influence L2 speakers' interpretations (explicit choices), but online eye-movements show that the L2 speakers *are* able to compensate for assimilation, similar to L1 speakers: Both groups show more looks to the unassimilated form in fast-rate contexts. However, (i) the differences in L1 and L2 speakers' offline choices and (ii) timing differences in eye-movements (L2'ers showing the effect *later* than L1'ers) demonstrate that even in high-proficiency L2'ers, there is still an intrusion of the L1 phonological system into L2 processing, indicating interference between the two systems. When no other phonological systems interfere (L1 English speakers), we find clear effects of compensation-for-coarticulation in real-time processing.

Capturing psycholinguistic processing effects using Amazon Mechanical Turk

Kelly Enochson & Jennifer Culbertson (George Mason University)
kenochso@gmu.edu

Typical psycholinguistic research in the lab requires substantial resources—access to a sizeable population of undergraduate students, infrastructure for recruiting, and means of compensation. Not only can completing experiments be expensive and time consuming, but the generalizability of any findings may depend on features of the sample, like education level and age. Researchers in psychology and cognitive science have recently begun looking to online services as an inexpensive and fast alternative (Crump, McDonnell & Gureckis 2013). Amazon Mechanical Turk (MTurk) is one such service: an online crowd-sourcing marketplace that allows researchers to post experiments, and automatically recruit and compensate participants from disparate demographic groups. Here, we show that MTurk provides a promising framework for research in psycholinguistics—including when accurate measurements are crucial.

The current study uses recently developed software, ScriptingRT, which captures response times accurately over the web using Flash (Schubert, Murteira, Collins & Lopes 2013). The goal is to replicate three robust psycholinguistic effects to demonstrate the validity of using MTurk for response time research. The findings we attempt to replicate are, (i) the reduction in processing cost associated with reading a pronoun versus a DP, (ii) the filler-gap effect, and (iii) agreement attraction. The first finding has been tied to definiteness, which a number of studies have shown to impact processing and acceptability judgments of referring expressions in various linguistic contexts. Self-paced reading experiments have shown faster reading of pronouns compared to definite DPs (Warren & Gibson 2002), and definite compared to indefinite DPs (Hofmeister & Sag 2010). The second finding concerns filler-gap constructions, in which there is a grammatical dependency between a fronted element and its original syntactic position. Many studies have demonstrated a processing cost associated with a filler-gap dependency (Wanner & Marastos 1978, Stowe 1986, Crain & Fodor 1985, Clifton & Frazier 1988, Boland, et al., 1995). The third finding is agreement attraction, which occurs when a verb spuriously agrees with some nearby plural DP rather than with its singular subject. Agreement attraction effects have been found in a number of constructions, including with prepositional phrase modifiers (Bock & Miller 1991, Pearlmutter et al. 1999), relative clauses (Bock & Miller 1991, Wagers et al, 2009), and *wh*-fronted constructions (Badecker, MS).

The current study used ScriptingRT to create a Flash movie instantiating a self-paced reading task (King & Just 1991). The Flash movie was then embedded in an html page and posted on MTurk. 34 participants were recruited and compensated ($1 each) via MTurk. The task was comprised of 48 test and 72 filler items (randomized using functionality available in ScriptingRT). Test items were grammatical *wh*-questions (e.g., “Which cars has the salesman found easy to sell?”), Pronoun/DP status of the subject, subject number, and *wh*-element number were manipulated, allowing us to test all three effects of interest.

The results revealed significantly faster reading times for pronouns compared to DP subjects (β=-.028±.009, p=.002), successfully replicating this robust effect. Results also revealed a significant slow down following the verb (β=.397±.016, p<.001), replicating the filler-gap effect. By contrast, no significant increase in reading time was observed in sentences that mismatch in subject and *wh*- number, as would have been indicative of agreement attraction. One possible explanation for our failure to replicate this finding is suggested by Pearlmutter et al. (1999): the magnitude of attraction effects in grammatical sentences may be smaller than ungrammatical sentences. Ongoing studies are testing agreement attraction effects over MTurk in two well-attested contexts—with prepositional phrase modifiers and with relative clauses, including ungrammatical test items. In general, however, our results demonstrate that web-based experiments can effectively capture psycholinguistics effects, offering a relatively low-resource alternative to lab-based studies.

Pupillometry – the Index of Cognitive Activity as a measure of linguistic processing difficulty

Vera Demberg & Asad Sayeed (Saarland University)
vera@coli.uni-saarland.de

Pupil size has long been known to be affected by cognitive load. We here explore, in a series of experiments, the use of a recent variant of pupillometry, called the Index of Cognitive Activity (ICA; Marshall 2000) for robustly measuring linguistic processing difficulty. The ICA quantifies cognitive load through the *frequency of small and rapid changes* in pupil size. An advantage of this method in comparison to overall pupil size is that the ICA is robust with respect to effects of changes in lighting and movements of the eye balls (it filters out slow and larger effects on pupil size), and that it is a more dynamic measure (much lower self-correlation in time makes it possible to more easily separate the effects of overlapping stimuli).

ICA in language comprehension. The present experiments are the first to investigate whether linguistic processing difficulty can be measured using the ICA measure. In three word-by-word self-paced reading experiments with simultaneous eye-tracking (each 24 items, 24 subjects), we tested (1) locally ambiguous German subject versus object relative clauses, (2) semantic anomalies, (3) German grammatical gender violations. We found significant effects of our linguistic manipulations for each of the three experiments in the left eye's ICA (using mixed effects models regression analysis with random slopes of condition under item and participants in all analyses; (1) $t=2.17$, $p<0.05$; (2) $t=2.47$, $p<0.05$; (3) $t=2.16$; $p<0.05$) during the time period of 300msec till 1200msec after the critical word first appeared on the screen.

ICA in single task steering. In a single task steering experiment (a standard tracking task cast into the driving scenario), we also found a highly significant cross-correlation (peak correlation at a lag of ca. 800 msec) between steering velocity (i.e. the speed with which the user turns the steering wheel) and the ICA, with more frequent rapid dilations of the pupil when the user makes a fast steering movement.

ICA in dual task: language comprehension and driving. After these first validation experiments, we ran all three linguistic experiments using spoken stimuli in a dual task setting (each 24 items, 24 subjects) with a simultaneous car steering task, and were able to replicate the finding of a significant effect in the critical regions (700mec till 1300msec after critical region speech onset) in the ICA of the left eye (1) $t=2.32$, $p<0.05$; (2) $t=2.19$, $p<0.05$; (3) $t=2.06$, $p<0.05$. We additionally found significant effects of the steering task in the ICA in both eyes.

ICA in the visual world paradigm (VWP). Our final experiment collects ICA in a VW experiment with spoken language stimuli that differed in the use of a causal versus concessive discourse marker. We found significantly higher ICA on both eyes for the concessive compared to the causal connector ($t=2.26$, $p<0.05$ (left); $t=2.14$, $p<0.05$ (right)), consistent with a P600 effect on this region in a related ERP experiment. These results are promising for the ICA to be used in the VWP for assessing processing difficulty in addition to visual attention.

Conclusions. We believe that the ICA method has great potential for measuring linguistically induced processing difficulty in situated dual task experiments and the VWP. The fact that we found significant effects in the expected direction for such different tasks as steering and language comprehension also indicates that the method might be applicable to a wide range of tasks, providing a measure of demand on cognitive control processes. The ICA merits more investigation, in particular with respect to hemispheric differences, and neural pathways.

References:

Marshall (2000). U.S. patent 6090051

Relativized Minimality: A systematic investigation on intervention effects

Sandra Villata (University of Geneva), Luigi Rizzi (University of Geneva & University of Siena), Akira Omaki (Johns Hopkins University) & Julie Franck (University of Geneva)
sandra.villata@unige.ch

Building on featural Relativized Minimality (fRM, Rizzi 2001), Friedmann et al. (2009) hypothesize that a local relation between an extracted element and its trace is disrupted by an intervening element as a function of the feature overlap between the intervener and the extracted element: interference should be (i) maximal whenever the featural specification of the intervener completely overlaps the featural specification of the extracted element (feature identity), and (ii) moderate whenever the featural specification of the intervener only partially overlaps the featural specification of the extracted element (feature inclusion). However, prediction (i) was questioned by the observation that the extraction of a Discourse-linked (Pesetsky, 1987) wh-phrase over another D-linked wh-phrase is acceptable (e.g., *Which problem* did you wonder *which student* solved?), more so than the extraction of a bare wh-phrase over another bare wh-phrase (e.g., *What* did you wonder *who* solved?). Featural RM also predicts that a configuration of feature inclusion in such cases of extraction (e.g., *Which problem* did you wonder *who* solved?) should give rise to higher acceptability rates than both D-linked and Bare Identity. We explored the predictions of fRM with three acceptability judgement experiments on French indirect questions in a 2x2x2 design manipulating: (a) Extraction of a wh-phrase over another wh-phrase (Extraction vs. No Extraction), (b) D-linking of the first wh-phrase, Wh1 (D-linked vs. Non D-linked), and (c) D-linking of the second wh-phrase, Wh2 (D-linked vs. Non D-linked).

Experiment 1 tested 40 participants' acceptability judgements of 32 sentences on a 7-point Likert scale. Experiment 2 tested 20 participants on 320 sentences on a binary scale. Experiment 3 tested 49 participants with the same materials and procedure as Experiment 1 except that sentences were preceded by a context story, so as to investigate the role of context in the processing of D-linking. Participants' ratings were *z*-score transformed and in order to partial out any influence from variables other than extraction in the assessment of the role of D-linking in extraction, we corrected the ratings of sentences with extraction by the ratings of the corresponding sentences without extraction (e.g., *Who* wonders *who* solved this problem?).

Results from the three experiments are highly consistent. A main effect of extraction was found, with dramatic sentence degradation in sentences with extraction. A main effect of Wh1 revealed that acceptability scores improve when Wh1 is D-linked. Tests on interactions showed a significant interaction between Wh1 and Wh2 attesting that when Wh1 is D-linked, higher ratings were found when Wh2 is also D-linked, while no effect of Wh2 was observed when Wh1 was non D-linked. No effect of context was found, suggesting that the ameliorative effect of D-linking in the extraction condition is due to the presence of a lexical restriction rather than an interpretive property like D-linking (as in Friedmann et al., 2009). Finally, comparisons on the individual structures corrected for baseline revealed that D-linked Identity is the most acceptable condition, followed by Inclusion, which is in turns more acceptable than Bare Identity and Inverse Inclusion, which were similarly rated. In the discussion, we address the consequences of these results for different approaches to intervention effects, and the potential relevance of content-addressable models of memory for language comprehension (e.g., McElree et al., 2003).

References

Friedmann, N., Belletti, A., Rizzi, L. (2009). *Lingua, 119*, 67-88.
McElree, B., Foraker, S., Dyer, L. (2003). *Journal of Memory and Language, 48*, 67-91.
Pesetsky, D. (1987). MIT Press.
Rizzi, L. (2001). In M. Baltin & C. Collins (Eds.), *Handbook of Contemporary Syntactic Theory*, Blackwell, Oxford, 89-110.

ERP indices of predictive processing

Edith Kaan, Evan Carlisle, George Collins, Nicholas Feroce & Amalia Reyes (University of Florida)
kaan@ufl.edu

There is ample evidence that listeners and readers make predictions as to what comes next in the linguistic input. Most of this research is based on visual world paradigms, which are naturally limited to the auditory modality and to the use of concrete nouns. Studies testing predictive processing using ERPs have mainly used paradigms in which the critical word forms an expected or unexpected continuation of the preceding context, or in which the degree of expectancy is manipulated (for overviews, see Kutas, DeLong, & Smith, 2011; Van Petten & Luka, 2012). ERPs at the critical word in these paradigms therefore reflect processes related to the integration of the word, or detection and repair of a violation, rather than anticipation. In the present study, we introduce an ERP paradigm that allows us to investigate predictive processing in written language before the critical word. This paradigm involves the occasional lengthening of the interstimulus interval (ISI) (cf., Besson, Faita, Czternasty, & Kutas, 1997). In Experiment 1, we tested this paradigm by presenting participants with sequences of letters. Each letter was presented for 300 ms with a 200 ms blank screen between the letters. The letter sequences were either highly predictive (‘a’, ‘b’ , ‘c’, ‘d’), or random (‘s’, ‘j’, ‘m’, ‘d’). Occasionally a letter was followed by a 500 ms blank screen. A comparison of ERPs between the conditions at the lengthened blank screen showed a centrally distributed negativity for the predictive versus random sequences, which we interpret as a Contingent Negative Variation (CNV); in addition, an increased P2 was seen for sequential vs. random letters following a delay, but not in the no-delay conditions, suggesting that fewer attentional resources are needed to recognize the predicted letter after a delay.

Experiment 2 tested the delay paradigm in a sentence context. Sentences formed either high or low-cloze contexts for the same critical nouns. Critical nouns were high-cloze continuations of the high-cloze context (e.g., “The hairdresser cut my hair but...”), or plausible continuations of the low-cloze context (“I’ve never liked my hair but…”). The critical noun was occasionally preceded by a lengthened ISI (500ms rather than 200ms). The ISI manipulation had an effect only in the first half of the study. First, a larger posterior negativity was seen for high vs. low-cloze contexts during the delay, which was comparable to the CNV in Experiment 1. Second, the P2 at the critical noun was larger after a delay compared to no delay, but did not show any difference between high and low-cloze contexts. This suggests that a delay eases the perception of a next word, regardless of its predictability. Finally, the N400 at the noun was smaller in the high-cloze versus low-cloze contexts, especially in the delay conditions. This suggests that the delay strengthens the expectation for the upcoming word in the high-cloze contexts, and facilitates the lexical activation and semantic integration of the critical noun.

The above findings suggest that the delay paradigm can be successfully applied to the study of predictive processing during written sentence comprehension. A CNV was observed before the critical word when the presentation of this word was slightly delayed. This component is sensitive to contextual constraints, which allows for a more direct measure of predictive processing, separate from the integration of the expected or unexpected word itself.

References

Besson, M., Faita, F., Czternasty, C.& Kutas, M. (1997). What's in a pause: event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, *46*, 3-23.

Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190-207). New York, NY: Oxford University Press.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*, 176-190.

Effects of morphological and prosodic focus cues on topic maintenance in Korean

Kitaek Kim, Theres Grüter, & Amy J. Schafer (University of Hawai'i at Mānoa)
kitaek@hawaii.edu

Both *topicality* and contrastive *focus* increase referents' prominence and guide reference interpretation; yet their interaction is not fully understood (Arnold, 1998, 1999; Cowles et al., 2007; Kaiser, 2011). There are two opposing hypotheses as to *how* focus may affect choices about subsequent topics (or focus): Focus makes an argument more salient in memory (Birch & Garnsey, 1995), and thus a more likely topic in subsequent discourse (henceforth, H1). On the other hand, contrastive focus indicates the presence of alternatives relevant for interpretation (Rooth, 1992), thus potentially promoting maintenance of the current topic together with selection of a contrastive alternative as focus (henceforth, H2). The current study manipulates both prosodic and morphological markings of information structure in spoken Korean to test these hypotheses, using stimuli describing transfer-of-possession events as in (1)-(4).

The Korean topic/focus marker *–nun* is considered a topic marker on NPs in sentence-initial position carrying broad focus prosody (see 2), and a focus marker on NPs in non-initial position, where it usually appears with contrastive focus prosody (see 4; Sohn, 1999; Kim et al., 2013). It can also carry contrastive focus prosody in initial position (3). These three aural context conditions, plus a baseline condition (1) with broad prosodic focus and Nominative-marking on the Source were tested in a story continuation task (Rohde et al., 2006). 32 native Korean participants heard a context and then typed a continuation. The critical measure was whether they would maintain the Source argument ("Mr. C") as the topic of the subsequent sentence.

Broad focus: 1. Chelswu-ssi-*ka* Yengswu-ssi-eykey album-ul kenney-ss-eyo.
Chelswu-HON-NOM Yengswu-HON-DAT album-ACC hand-PAST-SES
2. Chelswu-ssi-*nun*…
Chelswu-HON-TOP…

Focus on *–nun*: 3. **CHELSWU-SSI-*NUN***…
CHELSWU-HON-TOP…
4. Chelswu-ssi-ka **YENGSWU-SSI-EYKEY-*NUN*** …
Chelswu-HON-NOM YENGSWU-HON-DAT-CONFOC …

All: 'Mr. C_{SOURCE} handed an album to Mr. Y_{GOAL}'

H1 predicts that a focused argument would be the subsequent topic and subject, i.e., more Source subjects in Cond3 and Goal subjects in Cond4 (vs. 2 or 1). H2 predicts that a focused argument activates a contrastive, alternative focus, thus increasing Alt(ernative)-Source subjects in Cond3 (e.g., *Mr. Alt didn't hand an album to Mr. Y*) and, in Cond4, both Original-Source subjects with an Alt-Goal dative (*Mr. C didn't hand an album to Mr. Alt*) and Alt-Goal subjects (e.g., *Mr. Alt didn't receive an album from Mr. C*).

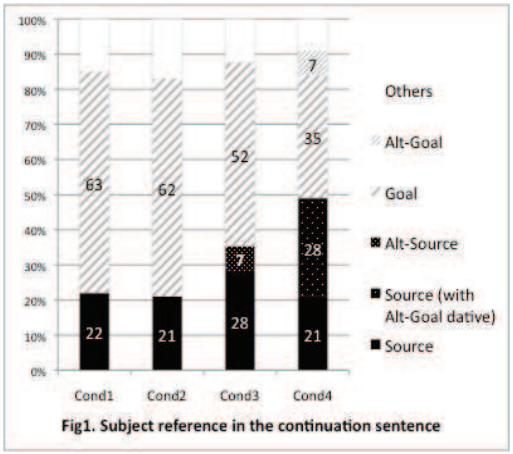


Fig1. Subject reference in the continuation sentence

Results (Fig1) for Cond1-2 replicate non-effects of –NOM vs. –TOP with broad focus (Kim et al., 2013). Cond4 shows increased Source-subject continuations (49%), often with Alt-Goal datives (28%), indicating that when morphological and prosodic focus cues coincide on the Goal, maintenance of the current, prosodically *less salient* (Source) topic is promoted (=H2). Cond3 shows increases in both Original-Source (28%) and Alt-Source (7%) subject continuations, suggesting that when a referent is simultaneously focused (prosody) AND marked as topic (*-nun* on initial NP), what is promoted is maintenance of the topic *qua* the set of entities including the referent and its alternatives (Féry & Krifka, 2008).

Determining necessary and possible values: An online study of modals and superlative modifiers

Yaron McNabb & Doris Penka (University of Konstanz)
yaron.mcnabb@uni-konstanz.de

There has been a great deal of work recently on the semantics and pragmatics of superlative modifiers like *at least* and *at most*. A particular challenge comes from the fact that these expressions give rise to ignorance inferences, which can be suppressed when certain superlative modifiers co-occur with certain modals (Geurts & Nouwen 2007). For example, *at least 50 minutes* in (1) implies that the speaker is unsure about the baking time and considers both exactly 50 minutes and higher numbers possible. When *at least* co-occurs with the necessity modal *has to*, as in (2), the authoritative reading enables the comprehender to understand that *50 minutes* specifies the lower bound of the range of allowed values, i.e. that 50 minutes and longer baking times are allowable, but not baking times shorter than 50 minutes.

(1) The cake baked for **at least** 50 minutes. (2) The cake **has to** be baked for **at least** 50 minutes.

The main purpose of this study is to determine which superlative modifier-deontic modal combinations are correlated with which readings (i.e. whether the modified numeral indicates the upper- or lower-bound of permissible values) as well as the time-course of the interpretation of these expressions in order to better adjudicate between competing analyses of these expressions, which make different predictions regarding available readings (G&N 2007, Büring 2007, Nouwen 2010, Coppock & Brochhagen 2013).

We conducted an incremental self-paced reading experiment, in which 40 German speakers read scenarios that included two interlocutors, A and B, followed by an UTTERANCE (3a) of A, which included a necessity (e.g. *muss* 'must') or possibility (e.g. *darf* 'can') deontic modal and a superlative modifier (*mindestens* 'at least' or *höchstens* 'at most'). The utterance was then followed by a DESCRIPTION sentence (3b), in which the number was either lower (UNDER CONDITION) or higher (OVER CONDITION) than the one used in A's utterance. Then, participants were asked whether the description was in accordance with the utterance.

(3) a. „Der Kuchen | {darf/muss} | {mindestens / höchstens} 50 Minuten | im Ofen | backen."|
"The cake | {can/has to} | bake | for {at least / at most} 50 minutes | in the oven."

b. John | bäckt | den Kuchen | {47 / 53} Minuten lang. |
John | baked | the cake | for {47 / 53} minutes. |

We found that the modal-superlative modifier combinations that led to the clearest interpretations were *at most* with either *must* or *can*, in which the modified numeral specified the upper bound, and *must* + *at least*, in which the modified numeral specified the lower bound. In *can* + *at least* participants tended to choose the lower-bound reading, but in the under condition, in 27.27% of the cases, values lower than the ones specified by the modified numeral (e.g. *47 minutes* in (3)) were accepted. Moreover, the reading times (in msec) in the under condition leading to a No response for the EVALUATION REGION—i.e. the numeral in the description sentences where we hypothesize participants arrived at a decision, e.g. *for 47 minutes*—were longer (M=1501.25) in *can* + *at least* than for those leading to a No response in *must* + *at least* (M=897.49; $p<0.05$). Reading times in the under conditions leading to a Yes response were longer (M=1607.24) in *must* + *at most* than in *can* + *at most* (M=1059.59; $p<0.05$), suggesting that the former is more difficult to interpret despite the clear response pattern.

Discrepancy Condition	*must + at least*	*can + at least*	*must + at most*	*can + at most*
Over	92%	89.23%	1.43%	4.29%
Under	0.01%	29.58%	89.19%	96.97%

Table 1: Means of "Yes" Responses

The response pattern and the reading times, taken together, suggest that *must* + *at least* and either modal with *at most* lead to clear interpretations as specifying the lower- and upper-bound, respectively, whereas *can* + *at least*, while showing a preferred reading, allows for two different interpretations, which come with a processing cost. The results of this study shed light on the on-line interpretation of sentences with deontic modals and superlative modifiers and call into question the current competing semantic and pragmatic analyses of these expressions, as none of them can account for the data presented in this study.

Generating implicatures from English NPs of the form *a/an X*: Generalized or local?

Li-Hsin Ning & Marina Terkourafi (University of Illinois at Urbana-Champaign)
ning3@illinois.edu

This study examines the direction of inferences from English indefinite noun phrases of the form *a/an X* and seeks to determine whether the resulting interpretations are generated by default or depend on the interaction of lexical semantics and local context. Neo-Gricean accounts make two different predictions in this regard. For alienable entities as in (1), the Q-heuristic favors the less specific interpretation (the specific interpretation of the indefinite NP is *not* the case), whereas for inalienable entities as in (2), the I-heuristic favors the more specific interpretation (strengthening to the stereotype).

(1) Horn's (1984) Q-principle and Levinson's (2000) Q-heuristic:
X is meeting a woman this evening. Q+> X is not meeting his wife, mother, sister etc..

(2) Horn's R-principle and Levinson's I-heuristic: stereotype
He broke a finger last night. R/I +> He broke his own finger last night.

However, if GCIs are indeed generated by default given a lexical item alone, then the indefinite article represents a problem for the Neo-Gricean account of GCIs. The two heuristics, Q- vs. I- (or Horn's R), yield opposing results and it is unclear which of the two interpretations, if any, is generated by default. In addition, recent experimental studies have called into question the division of GCIs into Q-, I-, and M- inferences proposed on theoretical grounds (Doran & Baker, 2009; Larson et al., 2009). These considerations motivated us to test experimentally the potential factors that may influence the interpretation of indefinite NPs of the form *a/an X* in English.

The lexical semantic factors manipulated include ALIENABILITY (+/-AL), ANIMACY (+/-AN) and CARDINALITY (1, 2, and 2+). Three experiments were administered to 14, 20, and 15 native speakers of English, respectively. First, a scenario norming task was conducted, in which 80 neutral scenarios were created for target noun phrases and normed to eliminate the possibility of contextual bias toward one of the possible interpretations (to avoid context-driven interpretations). Second, a scenario reading comprehension (forced-choice) task was carried out to test default interpretations of *a/an X*. The results showed that when the indefinite article is followed by an inalienable inanimate noun (*a waist*, *an arm* or *a finger*), I-enriched interpretations are generated, regardless of cardinality ($\chi^2(2)=30.309$, $p<0.001$). When the indefinite article is followed by an alienable noun of cardinality 1 (denoting entities with which the relationship in the real world is typically one-to-one such as *a physician* or *a university*), regardless of animacy, both Q-enriched and I-enriched interpretations are equally likely to be generated. On the other hand, when the cardinality is not 1 (nouns denoting entities with which the relationship in the real world is typically one-to-two or one-to-many such as *a professor*, *an earring* or *a book*), the indefinite article followed by an alienable noun regardless of animacy preferentially generates I-enriched interpretations. To examine the robustness of the Q- and I- inferences generated, we further conducted a rating experiment, in which we asked subjects to rate the acceptability of each of the two choices after reading a story. The results of the rating task support those of the reading task. When the indefinite article is followed by an inalienable inanimate noun (*a waist*, *an arm* or *a finger*), only I-enriched (more specific) interpretations are generated, regardless of cardinality ($F(1,262)=623.962$, $p<.0001$). As for alienable nouns regardless of animacy, when the cardinality is 2 or 2+ (*a professor*, *an earring* or *a book*), only I-enriched interpretations are generated; when the cardinality is 1 (*a physician* or *a university*), both I-enriched and Q-enriched interpretations could be generated in a given scenario. We conclude that the Neo-Gricean account of English NPs of the form *a/an X* can be upheld only for inalienable nouns. For all other noun classes, the relevant inferences are not attached to the construction but rather depend on the interaction of the lexical semantics of *X and* the local context. In other words, the inferences for alienable nouns can only be local. Our results provide partial empirical support for both default and local pragmatic accounts of conversational implicatures, suggesting that any predictions in this regard must be cast at a finer level of granularity than has previously been the case.

Influence of Visual Complexity on Referring Expression Generation

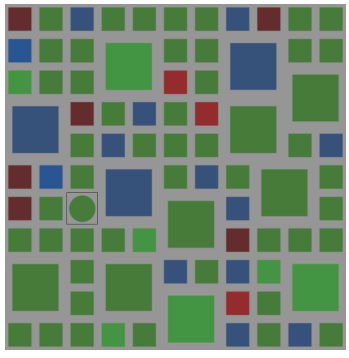
Hannah Rohde, Alasdair Clarke (University of Edinburgh), & Micha Elsner (OSU)
Hannah.Rohde@ed.ac.uk

In describing a target among distractors, speakers must extract relevant properties of a visual scene and formulate a coherent referring expression (RE) to pick out that target. Although linguistic cues influence how we see the world (Spivey et al., *Psych Science*, 2001) and properties of our visual system influence what we choose to say (Coco & Keller, *CogSci*, 2012), models of RE generation (REG) have largely sidestepped a role for vision (Clarke et al., *Frontiers in Psych,* 2013). Such models include the Incremental Algorithm (Pechmann, *Linguistics,* 1989; Dale & Reiter, *CogSci,* 1995). This algorithm successfully accounts for target overspecification in REG—i.e., speakers' inclusion of features of the target not required for minimal disambiguation. It does so by incrementally assessing target features (e.g., color, size, shape) and adding a feature to the RE if a serial check of distractors confirms that the feature excludes at least one distractor. The process terminates when the RE is unambiguous, though the speaker may start speaking earlier. However, the algorithm makes two large assumptions that are naive in light of findings from visual perception: (i) that the set of candidate features can be identified easily and (ii) that distractor checking is always a serial process.

Here we test how these assumptions hold up in visually complex scenes. A problem for (i) is that the set of potential features of a target is unbounded (beyond color/size/shape, why not consider the target's texture, orientation, lack of horns, etc.?), and more heterogeneous scenes may make the identification of relevant features harder. For (ii), the speaker can sometimes pre-attentively perceive a feature's effectiveness (e.g., shape excludes at least one distractor—in fact all—for a circle among homogenous squares). In visual search studies, the number of distractors slows viewers' response, but only if the distractors are heterogeneous. The Incremental Algorithm makes no prediction for such an interaction between set size and heterogeneity on REG, given that the role for vision is left unspecified. If findings from visual search extend to REG feature extraction/checking, then we predict this interaction in production.

Methods: 18 participants viewed 60 randomly generated stimuli consisting of a grid of colored shapes (see figure). The grid varied in set size (N = 25, 49, 81, 121) and heterogeneity (homogeneous, varying in color only, and varying in color/shape/intensity). Participants were instructed to describe the target so that a listener could quickly and accurately find that shape in the same grid. Because targets could be disambiguated by shape alone ("circle"), we also included 60 fillers with non-unique targets (which required a relative description: e.g., "leftmost red circle"). REs were coded for speech onset, offset, and the inclusion of relative descriptions.

Results: The crucial N x heterogeneity interaction emerged for post-onset speaking time: Little effect of N in homogeneous and color-only heterogeneous conditions (+5ms/object) but a large effect with more heterogeneity (+32ms/object). The lack of an interaction in onset times (and the tiny effect of N on onsets: +2ms/object) suggests that full feature extraction is non-trivial and is completed after speech onset. The post-onset pattern could arise if speakers provide more information (taking longer to communicate) and/or express the same information more slowly (more disfluency). Since a single word disambiguates all critical items, REs are overspecified if information is added. However, our index of overspecification (use of relative descriptions) does not vary across conditions. Full-scale disfluency annotation is under way.



Just as the Incremental Algorithm posits incrementality for feature inclusion in the RE, our results suggest that feature extraction/checking is likewise an active and ongoing process. To scale well, REG models should incorporate cues like scene complexity and be informed by findings from visual perception.

Lexical entrainment in deceptive interaction

Alessia Tosi, Holly Branigan, Adam Moore, & Martin Pickering (University of Edinburgh)
a.tosi@sms.ed.ac.uk

When people converse, they tend to converge on their use of referring expressions (*lexical entrainment*), a tendency so far studied in collaborative communicative contexts. We investigated how lexical entrainment is affected when one interlocutor tries to deceive the other.

Entrainment of lexical choices is one of many forms of verbal (and nonverbal) interpersonal adaptation (or alignment) occurring between people engaging in interaction. Shown to facilitate mutual understanding and the effectiveness of the communication, with positive repercussions on affiliation, trust, and liking between interlocutors (e.g., [1]), it remains controversial whether alignment occurs as the result of audience design [2] or more automatic cognitive processes (e.g., priming, cf. [3]). Deception has been shown to involve sophisticated socio-cognitive representations and skills (e.g.,self- and other-monitoring), and extra attentional and control resources associated with inhibition of a truth-bias and production of a conflicting response. As such, deceptive interactions might display different patterns of entrainment than the 'cooperative' interactions normally studied.

One possibility is that deceivers might show greater lexical entrainment than truth-tellers, in order to exploit the benefits of alignment found in cooperative interactions, e.g., increased sense of affiliation and gain of partner's trust. Alternatively, deceivers might show less entrainment than truth-tellers, either because they perceive themselves as distant from their dialogue partner, a form of disaffiliation that may result in decreased behavioral alignment, including lexical entrainment, or because the cognitive load induced by deception may impair deceivers' attentional resources, hence their depth of processing of linguistic input.

We used a picture-naming paradigm, in which participants (N=46) were primed by a confederate with highly acceptable-yet-disfavored names for pictured objects (e.g., "tumbler" for glass). Participant and confederate took turns naming objects for each other and choosing objects matching their partner's description. Every few trials, they did a memory test to identify objects not selected in previous matching turns; good performance depended on the correct objects having been previously named/selected. Participants in the **DECEPTIVE** condition were instructed to impair their partner's performance on the memory test by sometimes naming the wrong object (with the incentive of financial reward if they did so).

Mixed-effects analyses showed that deceivers were less likely to lexically entrain (52.5%) than truth-tellers (61.6%) ($p < .05$), and entrained increasingly less over the course of the interaction, whereas truth-tellers' entrainment remained stable over time ($p < .05$). Deceivers also performed worse in the memory test (17% vs. 42% correct) ($p < .0001$). Our results suggest that being deceptive disrupts the interactive process of lexical entrainment. Further investigation is needed in order to assess whether this reflects the deceivers' sense of estrangement from their partner, reflected in "linguistic distance", or the cognitive demands of deceptive behavior, whereby reduced attentional resources impairs deceivers' ability to engage in linguistic coordination. Deceivers' poorer performance on the memory tests, a task not directly related to the act of deceiving, provides support for the latter.

References

[1] Lakin, J.L., Chartrand, T.L., & Arkin, R.M. (2008). *Psychological Science,* 19(8), 816–822.
[2] Brennan, S.E., & Clark, H.H. (1996). *J. Exp. Psychol.-Learn. Mem.,* 22(6), 1482–1493.
[3] Pickering, M.J., & Garrod, S. (2004). *Behav. Brain Sci*. 27(2), 169–225.

Mechanisms of prosody production: Differences between children with and without ASD.

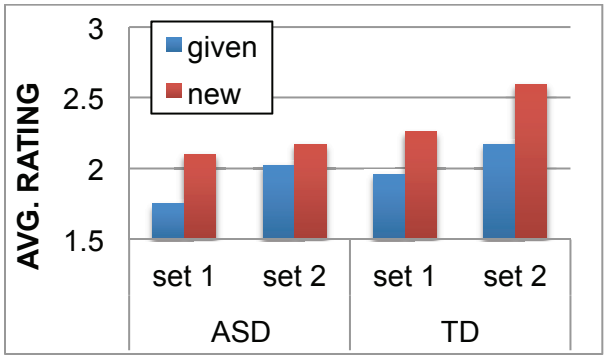
Jennifer E. Arnold, Elise C. Rosa, Mark Klinger, Patrick Powell, Alison Meyer (UNC Chapel Hill)
jarnold@email.unc.edu

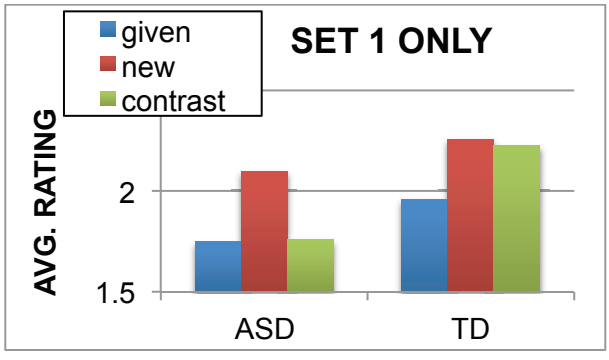
Autism spectrum disorder (ASD) is characterized by deficits in communication, frequently involving impairment in prosody. One prosodic function is to mark information status: speakers use prosodic prominence for New or Contrastive information, and reduced pronunciations for Given information. Several possible cognitive mechanisms may underlie this pattern (not mutually exclusive), any of which may be impaired in autism: 1) Rule use: Application of prosodic rules. 2) Production Facilitation: Given words are easier to produce than New words, supporting fluent, reduced pronunciations. 3) Audience design: Speakers choose pronunciations to facilitate comprehension for the addressee. We examined evidence for these mechanisms in ASD children and their typically developing (TD) peers.

Children ages 8-14 (17 ASD and 15 TD) performed a referential communication task with an experimenter. Both viewed a picture of four objects (e.g., penny, tiger, doorknob, grapes) and four shapes (triangle, circle, square, star) per trial on their own computers. The experimenter asked a question in one of three conditions: Given target (*Does the tiger go below the triangle?),* New (*Does the penny go below the grapes?*), or (in SET 1 only) Contrastive (*Do the tiger and the doorknob go below the grapes?).* The child instructed the experimenter to move an object based on a graphic cue, e.g. *No, the tiger goes below the grapes.* Two blind coders rated the acoustic prominence of the target word (**tiger**) on a six-point scale (1=reduced; 6= prominent). Additionally, we tested the child's sensitivity to the comprehension needs of the listener with a clarification filler half-way through the experiment (*Did you say dog or dock?).* Following this question, all scenes included two objects with similar names (i.e., cohorts). We compared pronunciations before and after this question as a measure of Audience Design.

As predicted, both groups used greater prominence for new than given targets. They also both used greater prominence following the clarification filler. This shows striking evidence for audience design, which has been contested for prosody (Bard et al., 2000; *JML*).

The Contrastive condition is important because it should elicit prosodic prominence through either Rule use or Audience Design. However, it involves the recent mention of the target word, which may lead to reduction via Facilitation. Critically, the groups responded differently to the Contrastive condition: for TD participants, the Contrastive condition was as prominent as the New condition. For ASD participants, it was reduced. This may reflect either a diminished representation of the conditions for appropriate prosodic prominence, or a stronger reliance on production facilitation over the addressee's needs.





On Questions and Speaker Ignorance

Xiaobei Zheng & Richard Breheny (University College London)
xiaobei.zheng.10@ucl.ac.uk

Previous studies [1] on the use of questions as a cue to speaker's knowledge state show that participants use these as a cue that the speaker refers to something known only to the participant. In this paper, we present studies exploring the basis of this result: Is it simply that the question form is associated with speaker ignorance, or are participants making mental-state inferences based on a richer informational basis. Here we manipulate the type of question while keeping the question form constant. In one condition, participants in a referential game based on [2] are asked Guess Question by the second participant, in a second condition, participants answer a Test Question that appears on the computer screen. In the first condition, the inference is that the asker is ignorant of the answer, in the second, the presumption is that the tester knows the answer. We show that in the Guess condition, participants perform better in terms of integrating information about speaker ignorance.

In an eye-tracking study, participants move objects around a 3*3 grid given instructions by a second player (a confederate). At the beginning of each trial, before the director gives the instruction, participants placed three privileged objects (not known to the director) in the three grid positions that the director cannot see. Then either the director (Guess question condition) or the program prompt (Test question condition) ask a question about the category of their privileged objects, in the form of a question (e.g. 'Are they fruit?'). Participants needed to judge the question through spoken sentences (Guess condition) or clicking on the screen choices (Test condition). As in similar visual-world studies, [2], test items ('Move the apple to the bottom middle') occurred when there was a competitor (Competitor condition) or an irrelevant object (Non-competitor condition) in privileged ground.

A second control experiment, examined whether simple verbal interactivity between the confederate and participants in Guess condition influenced participants' processing of the perspective information. Here, the confederate knew the categories of participants' privileged objects from the beginning in both conditions. The confederate either verbally assigned the categories and asked, 'ok?' (Interactive condition) or read the categories assigned by program, which appeared on the both screens (Non-interactive condition).

Participants' eye movements were recorded and analyzed in two 300ms windows after a baseline region (-200ms to 200ms from the onset of target word). In the first experiment there was a significant interaction between Question type and Competitor type during 500 to 800ms in by-items analysis and a trend in by-subject analysis, $F1(1,38)=3.796$, $p=0.059$, $F2(1,18)=4.770$, $p<0.05$. In the Test condition participants had larger target advantage scores in the Non-Competitor condition than the Competitor condition during 200 to 500ms, $F1(1,38)=4.107$, $p=0.05$, $F2(1,18)=4.275$, $p=0.053$, and also 500 to 800ms, $F1(1,38)=20.240$, $p<0.001$, $F2(1,18)=25.430$, $p<0.001$. However, in the Guess condition no difference was found between the competitor types. In the second experiment only a competitor type main effect was found during 500 to 800ms, $F1(1,65)=50.770$, $p<0.001$, $F2(1,36)=34.787$, $p<0.001$. No interaction between communicative type and competitor type was found during two time windows.

The results indicated that in the Test condition participants got larger interference from their privileged ground compared to the Guess condition, demonstrating that guess questions made participants more sensitive to the director's ignorance about privileged objects. As simple verbal interactivity had no effect in the control study, we conclude that it is the inferred mental state behind a question form that guides reference.

References

[1] Brown-Schmidt, S., Gunlogson, C., & Tanenhaus, M. K. (2008). Cognition, 107(3), 1122-1134.
[2] Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000).Psychological Science, 11(1), 32-38.

Pragmatic influences on the processing of ACD relative clauses

Edward Gibson (MIT), Pauline Jacobson (Brown), Peter Graff (Intel), Kyle Mahowald (MIT), Evelina Fedorenko (MIT), Steven T. Piantadosi (Rochester)
egibson@mit.edu

In a recent attempt to find on-line evidence for the existence of Quantifier Raising (QR), Hackl, Koster-Hale & Varvoutis (HKV; 2012, *Journal of Semantics*) reported that antecedent-contained-deletion (ACD) relative clauses modifying a quantified object noun phrase (NP; e.g., *every doctor… was*, as in (1a)) are easier to process than those modifying a definite object NP (e.g., *the doctor … was*). In contrast, in the full verb controls (as in (1b)), *every* has no advantage over *the*. HKV argue that this pattern of results supports a QR analysis of both ACD structures and quantified NPs in object position: under this account, both ACD resolution and quantified NPs in object position require movement of the object NP to a higher syntactic position. The advantage for quantified object NPs in ACD is hypothesized to arise from the fact that the quantified NP, but not the definite NP, has already undergone QR by the time ACD has to be resolved.

(1a) *The hospital was negotiating with every / the doctor that the organization was …*
(1b) *The hospital was negotiating with every / the doctor that the organization funded…*

We here challenge HKV's account. *First*, HKV's reading time data have two weaknesses: (i) HKV's account predicts not only an advantage for the "every" condition at the ellipsis site ("was"), but also a corresponding disadvantage at the object NP ("every doctor" vs. "the doctor"), because QR is hypothesized to have some cost, and this cost should simply occur at different time points across the two conditions. No such effect was observed. (ii) The reading time effect did not occur at the point where it was predicted (*was* vs. *funded* in (1)), but 2 words later (Expt 1) or 3 words later (Expt 2). Furthermore, re-analyses of HKV's data (available at http://hdl.handle.net/1721.1/76676) suggest that the effect was not significant even at these locations. An additional attempted replication of HKV's Expts 1 and 2, using 80 participants in each (more than HKV's experiments: 50 and 48) also demonstrated no reliable effect.

And *second*, although the effect in the acceptability rating study appears to be robust, HKV's interpretation of that effect is problematic. Importantly, HKV's design compares ACD with full verb structures such as (1b) where the overt verb does not have the same meaning as the understood verb in (1). When the same verb phrase is used in 1b as in the main clause (*was negotiating with*), the interaction that HKV observed remains, but a main effect of determiner is also present, such that the "the" conditions are rated as less acceptable than the "every" conditions. This pattern suggests that whatever factor is driving the interaction may also affect the conditions with no ellipsis. Critically, because the no-ellipsis condition does not require QR, the effect appears to be due to factors that have nothing to do with QR.

We present 6 rating experiments that provide evidence for an alternative, pragmatic, explanation of HKV's reported results: when the events in the two clauses are the same, there is pressure to use – instead of *the NP* – a competing form with *also* or *same* unless some prior connection can be established between the events (cf. Kaplan, 1984; Amsili, 2012). This pressure is absent for *every N* because the competing form is bad (**every same)* or – in the case of *also* – because *also* actually degrades the stimuli in the *every* condition. Our experiments show that the difference between the *the*-ACD and *every*-ACD conditions (a) persists even when the relative clause contains no ellipsis and thus nothing is forcing QR (Expts 1 and 2, using the materials from HKV's Expts 1 and 2, respectively); (b) disappears when either *also* (Expt 3) or *same* is included (Expt 4); and (c) disappears in supportive contexts (Expts 5 and 6), because the pressure to highlight the "sameness" disappears when the context establishes some other connection between the events (Kaplan, 1984). All of these effects are reliable in the rating data, using linear mixed-effects regressions, including participant and item slopes and intercepts (e.g., the context x determiner x ellipsis interaction effect in Expt 5, $t = 2.36$). None of these effects are predicted by the QR hypothesis, but they are predicted by our pragmatic hypothesis. Together, these findings show that HKV's QR hypothesis should be rejected in favor of a pragmatic account.

Processing canonical and non-canonical sentences in Turkish within a context

Barış Kahraman & Yuki Hirose (The University of Tokyo)
kahraman@phiz.c.u-tokyo.ac.jp

Previous studies have shown that canonical sentences are generally easier to process than non-canonical sentences. Keiser and Trueswell (2004) argued that old information tends to occur early, and new information tends to occur later in the sentences [1]. They showed that when the subject and object were mentioned in a supportive context, non-canonical sentences (OVS) were not more difficult to process than canonical ones (SVO) in Finnish. Keiser and Trueswell concluded that the processing difficulty of non-canonical sentences in the previous studies is due to violations of discourse factors, rather than the complexity/frequency based factors. However, in establishing discourse relationships, Keiser and Trueswell repeated nouns from the preceding contexts in their test sentences. Therefore, we are not sure whether their results are due to repeated noun benefit or givenness. We also want to test whether the impact of the givenness can be generalized to other languages having different word orders.

We conducted a self-paced reading experiment in Turkish (N=35) to examine the impact of the information status of the subject and object nouns on the processing of canonical (*SOV*) and non-canonical (*OSV*) sentences by manipulating the information status of the subject and the object. The target sentences were embedded in the beginning of a complement clause, and always preceded by a context such as “the ticket seller’s name was John.”, as shown in (1a)-(1d). To avoid unnaturalness and name repetition effects, we used pronouns when referring to given referents in the target sentences. The complexity/frequency based accounts predict that the verbs of canonical sentences should be read faster than the verbs of non-canonical sentences because canonical sentences have higher frequency and less complexity in Turkish (Demiral, 2007) [2]. On the other hand, discourse function based accounts predict that the processing difficulty of the verb will be affected by information status of the nouns [1].

We found only a main effect of information status of the nouns (F_1 (1,32) = 5.78, p=0.022; F_2 (1,21) = 5.51, p=0.029), with verbs after new-given noun order being read faster than the verbs after given-new noun order. This may be due to fact that pre-verbal position is focus position in Turkish, and given nouns are easier to be focused than the discourse-new nouns. In the first two regions, there was also a main effect of the information status of the nouns, and in the first region given nouns (pronouns) were read faster than the discourse-new nouns (F_1 (1,32) = 3.84, p=0.059; F_2 (1,21) = 5.33, p=0.031). In the second region, on the other hand, discourse-new nouns were read faster than the given nouns (F_1 (1,32) = 12.26, p=0.0001; F_2 (1,21) = 40.84, p=0.0001). This confirms that old-new information order is also easier to process than new-old information order in Turkish regardless of the surface word order, as in Finnish. Overall, this study suggests that non-canonical sentences are not always more difficult to process than the canonical ones, and that the processing difficulty of non-canonical sentences in the previous studies may indeed be due to violations of discourse demands. More importantly, complexity/frequency based accounts cannot explain the present results in Turkish, while discourse based factors can – thus extending Keiser and Trueswell’s results for Finnish.

(1a)	[He-NOM-$_{given}$	Mary-ACC-$_{new}$	cheats]	that the stationmaster	said.
(1b)	[Mary-NOM-$_{new}$	he-ACC-$_{given}$	cheats]	that the stationmaster	said.
(1c)	[He-ACC-$_{given}$	Mary-NOM-$_{new}$	cheats]	that the stationmaster	said.
(1d)	[Mary-ACC-$_{new}$	he-NOM-$_{given}$	cheats]	that the stationmaster	said.

References

[1] Kaiser, E. and Trueswell, J.C. (2004). The role of discourse context in the processing of a flexible word—order language. *Cognition, 94*, 113-147. [2] Demiral, Ş.B. (2007) *Incremental Argument Interpretation in Turkish Sentence Comprehension.* PhD dissertation. Leipzig: Max Planck Institute.

What you believe is what you get: The role of individual beliefs in sentence comprehension

Heeju Hwang (Stanford University)
heejuhwang@gmail.com

INTRODUCTION Previous research suggests that interpretations of sentences are often influenced by real-world knowledge (e.g. Ferreira et al. 2002, Sanford & Sturt, 2002). The so-called Depth-Charge sentences are an example. The sentence *'No head injury is too trivial to be ignored'* is often misinterpreted (head injuries should not be ignored), whereas the superficially similar sentence 'No missile is too small to be banned' is correctly interpreted (missiles should be banned). The sentences are (mis)understood as such because the common beliefs are that head injuries, however trivial, require attention and that missiles, however small, should be banned (e.g. Wason & Reich, 1979). In addition to the common beliefs and opinions, however, cognitive processes are often also influenced by individual beliefs and opinions as evidenced in confirmation bias. People tend to selectively gather or recall information that confirms their belief, or interpret ambiguous evidence as supporting their beliefs (e.g. Lord et al. 1979). The present experiments examine the role that individual beliefs play in sentence processing.

EXPERIMENT 1 In order to address the question, we constructed Depth-Charge sentences on controversial issues which Democrats and Republicans are not likely to agree on (1).

(1) No gun bill is too restrictive to be passed.

As in Wason and Reich (1979), the relation between the adjective and the verb was controlled to conform to *the more x the less y* in all sentences (e.g. the more restrictive the less it should be allowed). Similar to Wason and Reich, we asked participants (n=16) to decide whether a paraphrased sentence (e.g. gun bills should be passed) is consistent with the original sentence. If participants' own beliefs influence comprehension, we expect that Democrats and Republicans should interpret the sentences differently. In particular, given that a point of view is governed by the relationship between the noun and the verb (e.g. gun bill – pass, Wason & Reich, 1979), Republicans should be more likely to misinterpret (1) to mean that gun bills should *not* be passed. Participants' political views were measured based on 12 questions that were part of a national survey conducted by the Pew Research Center. The results show that Republicans were indeed more likely to misinterpret (1) to mean that gun bills should not be passed (57% vs. 33%), suggesting that people may use their own beliefs to guide interpretation of sentences. The fact that Democrats and Republicans did not show any difference in interpretations of garden path sentences used as fillers (e.g. While Anna dressed the baby played in the crib) suggests that the effect is not likely attributable to other factors (e.g. comprehension ability) than participants' political views. We also found an effect of verb. In general, participants were more likely to misinterpret sentences that contained a negative verb (e.g. prohibit) than those with an affirmative verb (e.g. pass) (64% vs. 32%). This is consistent with previous findings that negatives pose more difficulty in comprehension (e.g. Cornish & Wason 1970, MacDonald & Just 1989) (There was no interaction b/w verb and political views).

EXPERIMENT 2 In order to see whether participants (n=19) could also take speakers' beliefs into consideration, we embedded Depth-Charge sentences inside either 'a Republican said' or 'a Democrat said' clause (2). Consistent with the results of Exp1, we found effects of political view (50% vs. 32%), and verb (60% vs. 31%) (no interaction b/w verb and political views).

(2) A Republican / Democrat said that no gun bill was too restrictive to be passed.

CONCLUSION The results suggest that in addition to common beliefs, people may use their own or speaker-specific beliefs as a heuristic to obtain the meaning of difficult sentences. We suggest that individual beliefs might override local semantic analysis in the same way as common beliefs, as for controversial issues, individual beliefs dictate what is acceptable or not.

(*All statistics are significant at equal to or less than .05 level.)

Influencing persistence of meanings in entailment-obstructing grammatical environments

E. Allyn Smith (Université du Québec à Montréal) & Kathleen Currie Hall (University of British Columbia)
smith.eallyn@uqam.ca

This study tests three prominent hypotheses about why some meanings (such as presuppositions) are persistent in entailment-obstructing environments (**EOE**s) where assertions are not. We show that none of these hypotheses match native speaker perceptions in a set of offline tasks, and we use the results to propose other factors that could play a role in processing.

Background: In EOEs, **primary** entailments are obstructed such that what is inferred in (1) is not inferred in (2) because of the negation. But some kinds of (**secondary**) entailments are supposedly not obstructed by these same environments, such as presuppositions and Potts' conventional implicatures (CIs), making them persistent (also known as 'projective'). For example, The entailment that Lance is blind is inferred in both (3) & (4) despite the negation because it is a CI, which makes it secondary rather than primary as it was in (1).

(1) Lance is blind.	**Inferred**: Lance is blind. (primary, basic)
(2) Lance is not blind.	**Not inferred**: Lance is blind. (primary, EOE)
(3) Lance, who is blind, sings.	**Inferred**: Lance is blind; Lance sings. (secondary, basic)
(4) Lance, who is blind, doesn't sing.	**Inf.**: Lance is blind. **Not infer.**: Lance sings (sec., EOE)

Why are secondary meanings persistent in EOEs? Hypotheses: (i) because they must already be shared knowledge before they are uttered (Stalnaker 1973, 1974); (ii) only some are, and those that are are lexically specified rather than derived via alternatives, or else are neutralizable (Abbott 2006, Abusch 2010). These divide secondary meanings into the truly persistent (the presupposition of cleft sentences and anaphoric particles like *too*) and the only sometimes persistent (the presupposition of factive verbs like *know* and verbs with a preparatory state like *win*); or (iii) because they are *not-at-issue*, which occurs when they are not relevant to the Question Under Discussion (QUD), which reflects the conversation topic (Simons et al. 2010).

Experiments: We gathered naturalness ratings for both presuppositions and CIs (to test (i)), different kinds of presuppositions (to test (ii)), and not-at-issue and non-supplied contexts (to test (iii)). The not-at-issue contexts were normed to be sure the secondary meanings were deemed non-relevant. Participants read one of the four types of sentences in (1)-(4) (representative of the four conditions) with or without a QUD-defining discourse context depending on the experiment and then answered three questions of the form *How surprised would you be to learn...*: the target question (e.g.*...that Lance is blind?*) and 2 distracter questions. Inferred propositions were predicted to be less surprising than non-inferred propositions. In total, across all studies, we tested 6 kinds of secondary meanings (4 presuppositions: cleft, *know, win, the*, & 2 CIs: epithets, appositives), 3 environments (basic, EOE: negation, and EOE: antecedent of a conditional), 2 polarities of question, and 2 types of context (absent and not-at-issue).

Results and Discussion: A mixed effects ANOVA, post-hoc t-tests and subject & item analyses for the 100+ participants showed main effects of meaning type and enviroment but not context. We found significant persistence for *the* and appositives, mid-level persistence for *know* and *win*, and low persistence for the cleft and epithets. These results are problematic for all accounts, as (i) cannot account for the persistence of appositives, (ii) cannot account for the low persistence of clefts, and (iii) cannot account for the lack of effect of not-at-issue contextualizing. Furthermore, there are a number of interaction effects that cannot be explained by any existing account of presupposition. For example, taking pairs of similarly-persistent triggers like *win* and *know* and dividing their EOE contexts into negation v. antecedent of a conditional (AOC) subparts, we see that *know* is significantly more persistent with negation than in an AOC, while *win* is the reverse. We build on ideas in Simons 2001 to show how factors other than the QUD but that similarly involve computation of speaker intent can explain these results.

Reference as a side effect of discourse expectations: the case of focus particles

Israel de la Fuente & Barbara Hemforth (CNRS-LLF, Université Paris 7-Diderot)
israel.delafuente@linguist.univ-paris-diderot.fr, barbara.hemforth@linguist.univ-paris-diderot.fr

Research on the mechanisms that underlie discourse processing has shown that language comprehension is, to a significant extent, driven by expectations and that we use (extra)linguistic information incrementally to anticipate upcoming discourse. Evidence for this assumption comes, among others, from literature in reference resolution and production. In particular, studies on implicit causality (IC) have shown that there are lexically triggered preferences for specific types of continuations associated with the semantics of certain verbs that render a referent relatively more accessible in a comprehender's discourse representation making it the preferred antecedent of a pronoun in a subsequent *because*-clause (e.g. Koornneef & Van Berkum, 2006; Garvery & Caramazza, 1974; Hartshorne, 2013), as in (1).

(1) a. John admires Mary because … [SHE]
b. John fascinates Mary because … [HE]

Recent accounts on the interpretation of IC verbs argue that IC biases are due to specific discourse expectations triggered by missing causal content in combination with a general processing constraint to "avoid accommodation" (Bott & Solstad, in press). Following this proposal, we argue that the same kind of expectations triggered by IC verbs can be extended to the level of the proposition. Specifically, we investigate the role of the focus-sensitive particles *even* and *only* in the construction of discourse representations in English and how these can also trigger expectations about upcoming discourse. We argue that different predictions can be made here as a result of the interaction between the semantic content of the focus-sensitive particle, on the one hand, and that of the discourse connector, on the other hand, in the process of establishing coherence between both clauses (Kehler, 2002; Kehler & Rhode, 2013; inter alia), as in the examples (2)-(4).

(2) a. Even / b. Only John interrupted Mary when… [a. SHE – 83,7%; b. SHE – 84,9%]
(3) a. Even / b. Only John interrupted Mary because… [a. SHE – 64,5%; b. HE – 58,4%]
(4) a. Even / b. Only John interrupted Mary although… [a. HE – 67,8%; b. SHE – 50,5%]

Based on evidence on cleft constructions in French, German and Spanish, we predict that speakers expect temporal subordinates (2) to contribute to the backgrounded/presupposed part of the utterance. This should result in a dispreference for antecedents within the scope of the focus particle (e.g. John) (similar to the anti-focus effect proposed by Colonna et al., 2012). For causal and concessive relations (3-4), however, we assume that expectations depend on potentially missing or unspecified content. In (3a) *even* demands an explanation about the likeliness of the predicate that would explain why John, who is the least likely person to interrupt Mary, did so. In (3b) *only* seems to demand an explanation for the exhaustiveness of the entity in its scope, that is, why John and nobody else interrupted Mary. Following previous accounts on concessive subordinates (e.g. the concessive "q although p" implies the same counterfactual as that entailed by the causal "~q because p", König, 1991), we expect the opposite pattern in sentences like (4) to that elicited in (3).

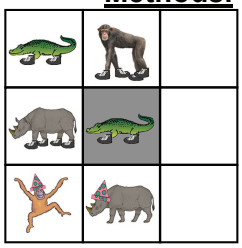
Our predictions were born out in a series of questionnaires in English where participants (n=147) were asked to provide an appropriate continuation to sentence onsets like those in (2)-(4) via Amazon's Mechanical Turk. Our results (given in mean percentages next to the examples (2)-(4) above) were corroborated by a series of log-linear mixed-effects model analyses with focus particle and discourse connector as fixed effects that yielded highly significant interactions of our two experimental factors. Crucially, these results were replicated in a series of parallel questionnaires in French, which supports the cross-linguistic validity of our claims.

Pragmatic narrowing in reference resolution: Domain restriction & perspective taking

Florian Schwarz (University of Pennsylvania) & Dan Grodner (Swarthmore College)
florians@ling.upenn.edu

Comprehenders have to juggle various types of contextual information, on top of the literal meaning expressed, such as that coming from Perspective Taking (PT) and Domain Restriction (DR). PT involves differentiating shared information from information that is only available to the listener (privileged ground). Some researchers have argued that comprehenders are initially ego-centric, failing to distinguish shared from privileged ground, and others that they are distinguished right away (Barr 2008; Brown-Schmidt & Hanna 2011). Domain restriction involves restricting the set of individuals under consideration: *The desk is covered with books* is normally understood to be implicitly restricted so as to pick out the contextually unique desk. DR can be seen as arising from a Question Under Discussion (Roberts 1996), which encodes the purpose of the ongoing exchange. The question of the processing time-course of DR is parallel to the case of PT: are incoming linguistic expressions initially evaluated literally, independently from contextual considerations, or are contextual constraints immediately applied? An additional question is whether or not DR and PT are different facets of the same underlying process of social coordination, or whether they draw, at least in part, on separate resources.

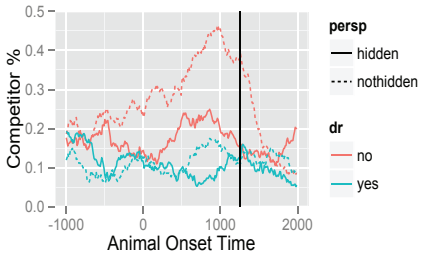
Methods. Subjects played a game where animal figures with accessories were moved about a 3x3 grid, with a goal of minimizing the number of adjacent animals with like accessories. A sample initial display is shown on the left. A confederate director provided instructions to a participant matcher. The matcher heard the instructions in another room via headphones. Privileged ground was indicated by shading parts of the subject's display gray (they were blocked on the director's screen). Most trials involved actual live reasoning by the director, but critical instructions were pre-recorded, as in (1).



No-DR/Hidden display from the matcher's perspective

(1) *Ok, let's start by dealing with the **boots**. Move the alligator* {**Hidden**: ∅ / **Not-Hidden**: *on the top left*} *to the bottom right.*

The competitor (the middle alligator) had the same accessory as the target in the **No-DR** condition, but a different one in the **DR** condition. Thus the context sentence potentially eliminated the competitor from consideration in the **DR** condition. The perspective conditions were **Hidden** vs. **Not-Hidden**. This resulted in a 2x2 interaction design.



Results: Analyses were carried out on the eye movements of 26 participants. Elogit-transformed proportions of looks to the competitor as well as target advantage scores were modeled with mixed effects regression. Both the **DR** and **PT** conditions exhibited significantly more looks to the target relative to the base-line (No-**DR**:**NotHidden**) within 200ms. Interestingly, there were significantly more looks to the competitor in the **No-DR**:**Hidden** condition than in the **DR**:**Not-Hidden** condition suggesting that DR may be more powerful than PT. Further, there were marginally fewer looks to the competitor in **DR-Hidden** versus **DR**:**Not-Hidden** 600-800ms after animal onset. Thus PT helped identify the target even when DR was operative.

Discussion: Immediately upon hearing *the alligator*, subjects do **not** go through a phase of indiscriminately looking at any visually present alligator. When the context narowed the DR to a specific accessory, attention is restricted to the alligator with that acessory. Similarly, if the competitor is in privileged ground, fixations are biased towards the shared ground alligator. Interestingly, the former effect is stronger than the latter, suggesting that the processes involved are not completely identical. The results argue against an initial egocentric processing stage and suggest that comprehension is immediately relativized to considerations about the state of the discourse, including the perspective of the speaker and the shared goal structure reflected in the Question Under Discussion that shapes DR.

Author Index

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Notes