The Ohio State University

Working Papers in Linguistics

No. 60

Edited by
Mary E. Beckman
Marivic Lesho
Judith Tonhauser
Tsz-Him Tsui

The Ohio State University
Department of Linguistics

222 Oxley Hall
1712 Neil Avenue
Columbus, Ohio 43210-1298 USA

Spring 2013

INTRODUCTION

This volume of the Ohio State Working Papers in Linguistics is a *varia* issue. It reflects a diversity of interests, containing papers that represent a good cross-section of the multiple sub-disciplines in the field of linguistics that are a focus of research in the department. Several authors even combine multiple sub-fields within the same paper. Rather than impose unwarranted interdisciplinary boundaries by attempting to group the papers by research area, we have instead ordered them simply by the alphabetical order of the (first) author's surname.

This volume is a landmark volume in two ways. First, it is volume 60, a special number by being five complete cycles of the Chinese zodiac. Second, it appears on the 50th anniversary of the first two degrees in linguistics to be awarded in the newly created Division of Linguistics at the Ohio State University, specifically the BA degree to Sandra A. Thompson and to Ray Saunders.

In June 2015, the department will celebrate the 50th anniversary of the establishment of doctoral program in linguistics at the Ohio State University. We are currently planning to publish, as volume 61 of *OSUWPL*, a special anniversary issue that will be a collection of papers by graduates of this program. This introduction also serves as a call for submissions to this volume. If you earned a Ph.D. from the Ohio State Department of Linguistics and would like to contribute a paper to the volume, please let us know by sending an e-mail to the editorial committee at:

wpl@ling.osu.edu

For subscription information and back issues, please visit us on the web at:

http://linguistics.osu.edu/research/publications/workingpapers

Ohio State University
Working Papers in Linguistics

No. 60

Table of Contents

MULTIPLE ANTECEDENT AGREEMENT AS SEMANTIC OR SYNTACTIC AGREEMENT[1]

Cynthia A. Johnson
Ohio State University

Abstract

In this paper, I challenge the argument put forth by Corbett (1991) that, within multiple antecedent agreement, the two possible agreement strategies, Resolution and Partial Agreement, can be viewed as semantic and syntactic agreement, respectively. Resolution, while semantically motivated and involving input from all of the agreement controllers, is not the same as semantic agreement in single-antecedent contexts. Partial Agreement, which relies on the morphological features of only one of the antecedents, still requires reference to the semantic features of both antecedents, as this strategy is more likely when the controllers are inanimate. Instead, I propose that the distribution of the two strategies – which nonetheless reflects the Agreement Hierarchy (Corbett 1979) and the Predicate Hierarchy (Comrie 1975) – is a product of the cognitive difficulty multiple antecedent agreement contexts pose for the speaker, such that the rules for this context are really part of broader principles within and across languages.

[1] An earlier version of this paper was presented at the 15th International Morphology Meeting in Vienna, Austria.

Cynthia A. Johnson. Multiple antecedent agreement as semantic or syntactic. *OSUWPL* Volume 60, Spring 2013, 1-9.

1. Introduction

Agreement with multiple antecedents provides surprising and interesting data for theories of syntax and agreement (Corbett 1991:261). In instances of multiple antecedent agreement, it is not immediately obvious from the features of the controllers alone what the features of the target should be. Unlike agreement with a single antecedent, where the controller and the target almost always share the same agreement features, agreement with multiple antecedents produces what Corbett (1991) refers to as an agreement mismatch between the controllers and the target(s).

Compare the following two examples from Latin. In example (1), agreement with a single antecedent, the controller is *Scipio*, a masculine singular noun. As expected, the targets share the same features as the controller: the verb *sit* agrees for the correct number, singular, and the adjective *clarus* agrees for both gender and number, masculine singular.

(1) Single Antecedent Agreement
sit Scipio clarus
be-3.SG Scipio.M.SG illustrious.M.SG
'Let Scipio be illustrious.' (Cicero, *Cat*. i*v.21*, from A&G)

On the other hand, in example (2), there are two antecedents (in the form of a conjoined noun phrase, *labor voluptasque* 'labor and desire') which control agreement on the target verb/past participle *esse iunctum*[2] 'be-bound.INF'. Since the target cannot be masculine (following *labor*) and feminine (following *voluptas*) at the same time, we have the precise context for an agreement mismatch: the features of the target will never match both controllers at the same time, since only one set of feature values can be expressed by a target. What we find instead is a small range of possible feature combinations for the target, modeled on either just one of the antecedents or according to a set of semantically-based rules.

(2) Context for Multiple Antecedent Agreement
Labor voluptasque ESSE IUNCTUM
labor.M.SG delight.F.SG.-and be bound
'Labor and delight are bound'

In Latin, and across languages more broadly, these possible feature combinations result from two main strategies for approaching multiple antecedent agreement: Resolution (3a) and Partial Agreement (3b) (terminology from Corbett 1991 and Wechsler & Zlatić 2003). Resolution appears to be more semantically motivated: the target's features are more or less "computed" by adding up the features of the controllers. As a result, the target is always plural (reflecting the semantic transparency of number agreement in this context) and the gender is determined according to language-specific

[2] This is not the usual citation form for Latin verbs (the first person singular present indicative is primarily used), but it represents a "neutral" form of the past passive verb: the auxiliary is in the infinitive (so as not to express gender) and the participle is in the neuter singular, the common citation form for adjectives.

rules, which often reference semantic features, e.g. animacy. Partial Agreement, on the other hand, does not display the same semantic motivation: rather than involving equal contribution of both of the controllers, the target's features match those of only one of the controllers. In Latin, the controller that serves as the basis for agreement is usually the one closest to the target (and so Partial Agreement in Latin is referred to as "Nearest Antecedent Agreement").

(3) Agreement Strategies

a. *Resolution*
formosi sunt verris et scrofa
handsome.M.PL are boar.M.SG and sow.F.SG
'The boar and the sow are handsome.' (Varro, *RR II.4.4*)

b. *Nearest Antecedent Agreement*
ut maxime **amicum** **cytisum** et medica
while very beneficial.N.SG snail-clover.N.SG and alfalfa.F.SG
'while snail-clover and alfalfa [are] very beneficial' (Varro, *RR II.2.19*)

One of the primary goals in investigating multiple antecedent agreement is to model the distribution of the strategies: within and across languages, in what contexts do we find Resolution and Partial Agreement? Corbett (1991) has proposed that the distribution of these strategies conforms to the Agreement Hierarchy (4a, Corbett 1979) and the Predicate Hierarchy (4b, Comrie 1975), if we view Resolution as semantic agreement and Partial Agreement as syntactic agreement.

(4) Typological Generalizations

a. *The Agreement Hierarchy*
attributive | predicate | relative pronoun | personal pronoun
⟵ syntactic agreement — semantic agreement ⟶
Nearest Antecedent Agreement — Resolution

b. *The Predicate Hierarchy*
verb | participle | adjective | noun
⟵ syntactic agreement — semantic agreement ⟶
Nearest Antecedent Agreement — Resolution

These hierarchies provide a basic typology of agreement: the more noun-like the target, the more likely we are to find agreement with the semantic features of the controller(s); the more verb-like the target, the more likely we are to find agreement with the syntactic (i.e. grammatical) features of the target.

Corbett's proposal, while supported by cross-linguistic research and single antecedent agreement structures, raises an important theoretical question: is Resolution really semantic agreement and is Partial Agreement really syntactic agreement? Is there independent evidence for viewing the strategies in this way, and should we expect multiple antecedent agreement to operate in the same way as the straightforward single antecedent agreement in (1)? In what follows, I provide a critique of Corbett's proposal

using data from Latin and suggest an alternative solution: one that takes into consideration linguistic performance. In sections 2 and 3, I discuss Resolution and semantic agreement, and Partial Agreement and syntactic agreement. In section 4, I present my performance-based view of agreement, which accounts for the distribution of strategies with reference to broader principles and rules within and across languages.

2. Resolution and semantic agreement

In Latin, there are two patterns of Resolution: in some instances, masculine is the resolved gender (5, repeated from (3a) above), but in other instances, neuter is the resolved gender (6).

(5) formosi sunt verris et scrofa
handsome.**M.PL** are boar.M.SG and sow.F.SG
'The boar and the sow are handsome.' (Varro, *RR II.4.4*)

(6) labor voluptasque … sunt iuncta
labor.M.SG delight.F.SG.-and are bound.**N.PL**
'labor and delight are bound...' (Livy. *AVC*, from A&G)

In both examples, there is a masculine singular antecedent and a feminine singular antecedent, i.e. the same grammatical features are present for both examples. The Resolution rules must therefore refer to a feature other than the grammatical gender of the antecedents to determine the resolved gender for each sentence.

The relevant feature is animacy, as discussed by the grammar handbooks (e.g. Allen & Greenough 1888) and previous literature (e.g. Corbett 1991), and supported by the data in my own corpus study (Johnson 2011). Animate antecedents use masculine as the resolved gender, while inanimate antecedents use neuter as the resolved gender. Given this connection between animacy and resolved gender, there is a clear semantic basis to these Resolution rules: at the very least, there is a connection between neuter grammatical gender and lacking biological gender and, perhaps to extend this connection, between masculine grammatical gender and having biological gender.

We have established that Resolution is semantically motivated, involving contribution of the semantic features of both antecedents, but is this actually semantic agreement? This would imply that the resolved gender follows naturally from the "adding up" of the controllers' gender features. Consider example (7) of semantic agreement in a **single** antecedent context.

(7) pars certare parati
part.F.SG to contend ready.M.PL
'a part [group of men] ready to contend'[3] (Vergil, *Aen*. v. 108)

[3]The larger context (7): laeto complerant litora coetu / uisuri Aeneadas, pars et certare parati.
'They filled the shores with a happy crowd / [some] to see the men of Aeneas, and a part ready to contend [in the games].' (Aeneid, v. 107-8)

In this example, the target *parati* is masculine plural, even though the controller, *pars*, is grammatically feminine singular. This is because *pars* refers to a group of men out in the world: the grammatical features of *parati* are modeled on the plurality and male-ness of the group of men that is the referent of *pars*.

The question we should consider is whether Resolution works in the same way: do the features of the target follow naturally from the "adding up" of the semantic features of the controllers? There are at least two problems here: first, there is necessarily a stipulated component to these rules. Resolution rules vary across languages: in Old Icelandic, for example, we find that all instances of Resolution are to the neuter gender, regardless of the semantic properties of the antecedents (Corbett 1991:80-3). If the gender of the group follows naturally from the semantics of the controllers, we would not expect to see this kind of variation. Additionally, the process of "adding up" genders to produce masculine or neuter gender is not semantically transparent. How does masculine and feminine combine to "equal" masculine or neuter gender? There is no transparent connection between the semantic genders of the controllers and the resulting resolved gender[4].

While Resolution is unquestionably semantically *motivated*, it is not the same kind of semantic agreement found in single antecedent agreement. Our explanation for the distribution of the two strategies should therefore reflect this fundamental difference between single antecedent agreement and multiple antecedent agreement.

3. Partial Agreement and syntactic agreement

The other multiple antecedent agreement strategy in Latin, Nearest Antecedent Agreement, occurs when the target shares the same feature values with only the closest controller, regardless of if the target follows (8a) or precedes (8b, repeated from (3b) above) the controllers. In cases where there is more than one agreement target, each target agrees with its closest controller (8c).

(8) a. Ibi Orgetorigis filia atque **unus** e filiis
There of-Orgetorix daughter.F.SG and one.**M.SG** from sons

captus est
was-captured.**M.SG**
'There the daughter and one of the sons of Orgetorix were captured.'
(Caesar, *BG 1.26*)

b. ut maxime **amicum** **cytisum** et medica
while very beneficial.**N.SG** snail-clover.**N.SG** and alfalfa.F.SG
'while snail-clover and alfalfa [are] very beneficial' (Varro, *RR II.2.19*)

[4] It was suggested (Corbett, p.c.) that the key to the semantics might lie in the Latin word(s) for 'group'; however, all three genders are represented by the various Latin 'group' words: *coetus* 'assembly' (m.), *classis* 'group, division' (f.), *decuria* 'gang, class' (f.), *conjectus* 'throwing together, i.e. collection' (m.), *collectio* 'collection' (f.), *corpus* 'body' (n.).

c. non **eadem** **alacritate** ac **studio** **quo**
not same.**F.SG** ardor.**F.SG** and zeal.**N.SG** which.**N.SG**
'[did not employ] the same ardor and zeal which [they had used to employ in land combat]' (Caesar BG. 4.24)

Again, we can ask a similar question: is Nearest Antecedent Agreement syntactic agreement? Syntactic agreement is defined as agreement consistent with the morphological features of the controller(s), without reference to the semantic features (Corbett 2006:156). In Nearest Antecedent Agreement, the target's features are only consistent with the morphological features of *one* of the controllers – and it is the controller that is nearer to the target. This local and linear dimension to Nearest Antecedent Agreement should not be ignored; as discussed below, Nearest Antecedent Agreement resembles a typical "ungrammatical" outcome of difficult long distance dependencies, i.e. attraction errors.

Additionally, while syntactic agreement implies no reference to the semantic features of the controllers, Corbett (1991) has stated that Resolution is more likely when the controllers are animate, i.e. there is some reference to a semantic feature when "choosing" which strategy to use.

Viewing Nearest Antecedent Agreement as syntactic agreement works only if this strategy is completely divorced from the semantics. In a broad sense, this is appropriate: there is no semantic rule that conditions the form of the target; it is only the proximity of one of the controllers that determines the features. However, we still need to explain why such a strategy does not involve input from all of the controllers (there could just as easily be a syntactic rule that computes the gender of the target from the morphological features of the controllers). Likewise, the semantic features of the controllers still influence the choice of strategy, even though the actual agreement process of Nearest Antecedent Agreement does not make reference to any semantic feature. This aspect of the distribution also requires explanation, as it is not explained by the hierarchies above.

4. Performance-based agreement: Gender assignment and Avoidance

If, on the basis of these facts, Resolution is not quite semantic agreement and Partial Agreement is not quite syntactic agreement, we need to explain why the hierarchies in (4) are still observed in Latin. In fact, even if Resolution **is** semantic agreement and Partial Agreement **is** syntactic agreement, such patterns still require explanation: the hierarchies in (4) are only typological tools that model common cross-linguistic patterns; by themselves, they offer no explanation as to why such patterns frequently occur. The solution I propose is one that takes into consideration linguistic performance, as evidenced in particular by the existence and acceptability of a strategy like Nearest Antecedent Agreement.

Unlike Resolution, Nearest Antecedent Agreement relies on linear and local relationships between the controllers and the target(s). As mentioned above, this strategy

– at least superficially – resembles what are typically referred to as attraction errors in other languages, e.g. the examples in (9) below.

(9) a. *Number Attraction*
The time for fun and games **are** over. (Bock & Miller 1991)

b. *Gender Attraction*
Stanze che sono anni e anni che sono
rooms.F.PL that be.3.PL years.M.PL and years M.PL that be.3.PL.

chiusi
closed. M.PL
'Rooms that have been closed for years and years' (Vigliocco & Franck 1999)

Along the same lines, Corbett (2006:170) has argued that, with respect to this agreement strategy, we should "perhaps be looking to psychologists, who have demonstrated the importance of first and last positions in lists in other domains."

There is also an inherent difficulty present in multiple antecedent agreement contexts. First, as discussed earlier, both strategies produce an agreement mismatch: the target cannot share the same features as both of the controllers. This makes the task of selecting agreement features more complicated than in single antecedent agreement contexts. Second, the gender system in Latin may be another source of difficulty for speakers. For animate nouns, the grammatical gender of the noun overlaps with the biological gender of the referent. This creates a gender system in Latin that in some instances references the natural sex of the referent, but in other instances, e.g. for inanimate nouns, it does not–it is purely grammatical, and thus has no relationship with the actual semantic properties of the referent.

Finally, multiple antecedent agreement is relatively rare: in my 300,000-word corpus study (Johnson 2011), there were only 47 unambiguous tokens, which means that speakers encounter this context far less than they do single antecedent agreement contexts. All of these facts about multiple antecedent agreement and Nearest Antecedent Agreement indicate – at least indirectly – that such a construction causes the speaker cognitive difficulty. The resulting strategies are a product of this difficulty, such that the strategies are a result of agreement done "on the fly," according to more general rules within and across languages.

4.1 Resolution as gender assignment

Resolution is simply gender assignment. Within Latin, both semantic and formal criteria are relevant for gender assignment: both the semantic features of the noun and the form of the ending are used for gender assignment, e.g. for borrowed words. In multiple antecedent agreement, the targets are all native words, so only the semantic criteria are used. In particular, it is the animacy value of the nouns that determines the assigned

gender. Wechsler & Zlatić (2003:182-3) have formalized this notion by proposing that coordinate noun phrases do not have an inherent lexical gender feature and so must be assigned a semantic gender based on a language-specific rule. In Latin, this rule – which operates not just in multiple antecedent contexts but elsewhere in the language – is one that simply correlates masculine grammatical gender with animacy and neuter grammatical gender with inanimacy. Since this rule applies in other feature assignment contexts, we are able to explain Resolution with reference to a broader rule within Latin.

4.2 Nearest Antecedent Agreement as Avoidance

Alternatively, rather than dealing with the complex problem of "adding up" genders via Resolution/gender assignment, the speaker can choose to avoid this problem altogether by simply agreeing with the closest antecedent. Nearest Antecedent Agreement is thus part of a larger category of what Hock (2007a) terms "Avoidance" strategies, whereby speakers employ a strategy that does not require them to produce a resolved form (where they must address the difficulty posed by the resulting agreement mismatch, the lack of semantic transparency in the gender marking, and the infrequency of the agreement context). Other Avoidance strategies found across languages include First Antecedent Agreement (e.g. in Slovene, Corbett 1991:266), restructuring the sentence completely (e.g. in Polish, Rothstein 1993), and gender neutralization (e.g. in German, Hock 2007b).

How does this performance-based view of multiple antecedent agreement fit with the hierarchies in (4) above? Rather than labeling Resolution as "semantic agreement" and Nearest Antecedent Agreement as "syntactic agreement," I account for the distribution of Resolution vs. Nearest Antecedent Agreement as one that is the product of the relative difficulty of different multiple antecedent agreement contexts. Resolution occurs when the semantics of the antecedents are more concrete (when the controllers are animate) and/or more relevant (when the target is more noun-like, i.e. when we must conceive of the coordinate noun phrase as a group).

Nearest Antecedent Agreement, on the other hand, is a product of the cognitive difficulty such contexts create, especially when the semantic features are less transparent (when the controllers are inanimate) and/or less relevant (when the target is more verb-like). The hierarchies are therefore explained not with reference to semantic or syntactic agreement – a problematic notion given the facts above – but according to the difficulty posed by multiple antecedent agreement more generally.

5. Conclusion

The Agreement Hierarchy and the Predicate Hierarchy are useful typological tools: in single antecedent agreement contexts, they accurately describe how likely a speaker is to agree with the semantic or morphological features of the controller. In multiple antecedent agreement, the same patterns are observed, provided we put Resolution on the semantic agreement end of the hierarchy and Partial Agreement on the syntactic agreement end. However, this conceptualization of the agreement strategies is met with significant theoretical problems: Resolution is not quite semantic agreement as defined in

single antecedent agreement contexts, and Partial Agreement still involves some reference to the semantic features of the controllers.

In order to account for the observed distribution of strategies, I instead propose that the patterns are a result of the overall cognitive difficulty associated with such an infrequent structure – a structure that is further complicated by the nature of gender marking in Latin and the agreement mismatch that necessarily results from multiple antecedent agreement. In this way, the strategies can be explained by broader principles within and across languages: Resolution as gender assignment, and Partial Agreement as a kind of Avoidance strategy.

References

Allen, Joseph Henry, and James Bradstreet Greenough. 1888. *New Latin grammar for schools and colleges.* Boston: Ginn and Company.

Bock, Kathryn, and Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23:45-93.

Comrie, Bernard. 1975. Polite plurals and predicate agreement. *Language* 51(2):406–18.

Corbett, Greville. 2006. *Agreement.* Cambridge, UK: Cambridge University Press.

Corbett, Greville. 1979. The agreement hierarchy. *Journal of Linguistics* 15(2):203-224.

Corbett, Greville. 1991. *Gender*. Cambridge, UK: Cambridge University Press.

Hock, Hans Henrich. 2007a. Agreeing to disagree: Agreement with non-agreeing antecedents, with focus on Sanskrit and Latin. East Coast Indo-European Conference, Yale University, June 2007.

Hock, Hans Henrich. 2007b. Early Germanic agreement with mixed-gender antecedents with focus on the history of German. Paper presented at the UCLA Indo-European Conference, 2-3 November 2007.

Johnson, Cynthia A. 2011. Multiple antecedent agreement in Latin. Paper submitted as the first qualifying paper for completion of doctoral program in Linguistics. Ohio State University Department of Linguistics.

Rothstein, Robert A. 1993. Polish. In Comrie, Bernard & Greville G. Corbett, eds. *The Slavonic languages*. London/New York: Routledge, 686-758.

Viggliocco, Gabriella, and Julie Franck. 1999. When sex and syntax go hand in hand: gender agreement in language production. *Journal of Memory and Language* 40:455-78.

Wechsler, Stephen, and Larisa Zlatić. 2003. *The many faces of agreement*. Stanford: CSLI Publications.

ON PHONICALLY BASED ANALOGY*

Brian D. Joseph
The Ohio State University

Abstract

In this paper I examine the role sound alone can play as the basis for analogical connections among forms, as opposed to more conventionally discussed factors such as paradigmatic structure, grammatical category, or meaning. Examples are presented here, mainly from English, that show sound effects in analogy at various levels of linguistic analysis, including phonetics, morphology, syntax, semantics, and the lexicon.

1 Introduction

Analogy, understood here in a broad sense to refer to any change in a given form due to the influence of another form, has a venerable history of study within linguistics, dating back to the Greek and Roman grammarians and their interest in the relationship between analogy and the origin of words and the origin of language itself. It is not surprising, therefore, that various textbooks on historical linguistics, perhaps most notably Anttila 1972/1989, have made clear the prominent role that analogy plays in the understanding of

* The material in this contribution is drawn from a presentation I have made in numerous venues since 2001 under various titles — too many to list — but beginning when I was a fellow at the Research Centre for Linguistic Typology at La Trobe University in July and August of 2001, at the kind invitation of R. M. W. Dixon and Alexandra Aikhenvald. I gratefully acknowledge the invaluable support of my residence there to this work, and thank the various audiences over the past few years who have contributed important insights to my thinking on the examples discussed herein.

Brian D. Joseph. On phonically conditioned analogy. *OSUWPL* Volume 60, Spring 2013, 11-20.

language change. Anttila's work, elaborated upon in Anttila 1977, established (perhaps, re-established) the semiotic underpinnings of analogical change.[1]

Still, even with so much attention to the topic, questions remain about analogy. One such question, given that analogy depends on a connection being made between two forms (the influencer and the influencee, so to speak), is just what sorts of connections can serve as the basis for analogical pressures and ultimately for analogical re-formations.

In this brief piece, I present a number of examples I have collected over the years that address this key question by demonstrating that one type of linkage between forms that must be recognized is a purely phonic one, based on sound alone. This is so even though sound is not generally thought of as a basis for analogical connections; most discussions of analogy in historical linguistics textbooks focus only on grammatical connections between forms, e.g. forms that are in the same paradigm (traditional "leveling" or "internal analogy") or forms that are members of the same grammatical category ("form class analogy" or "external analogy").

The general neglect[2] of a phonic basis for analogy is perhaps somewhat surprising, given that a phonic basis can be found in other aspects of language use. For instance, sound is critical in many types of language play, among them counting rhymes, such as *eeny, meeny, miny, mo* with its assonance and alliteration. Moreover, sound plays an important role, beyond simple rhyming patterns, in various sorts of literary expression; for instance, Miller 1982 has demonstrated complex phonic echoing within lines in Homeric epics, Watkins 1995 has shown the importance of phonic devices linked to thematic parallels throughout several ancient Indo-European poetic traditions, and Dawson 2005 draws attention to the effects of homoioteleuton, a phonically based poetic (and rhetorical) device, in the selection of certain dual and locative allomorphs in Vedic Sanskrit.[3] Further, even within recognized types of analogy, a phonic basis often is lurking. For instance, classic cases of 'contamination', which in one sense can be viewed as leveling within a 'semantic paradigm', can involve a phonic link. A relevant example is Late Latin *grevis*, which is generally believed to have developed from Classical Latin *gravis* "heavy" through 'contamination' with its semantic opposite *levis* "light", plus some influence likely from the semantically related (as a dimension adjective) *brevis* "short; brief"; however, even if the semantic links were important here — and I have no doubt that they were — there is a phonic link as well with *gravis/grevis*, *levis*, *brevis*, in

[1] Note also the excellent bibliography on analogy, Anttila & Brewer 1977, and various recent handbook-style treatments of analogy, especially Anttila 2003 and Hock 2003.

[2] There are exceptions; Vennemann 1972, for instance, with its discussion of 'phonetic analogy', clearly emphasizes that the notion of analogy must be extended to include connections made at the level of sound and not of grammar proper. Claims concerning the purely grammatical basis of analogy are to be found in work done within the framework of Optimality Theory, on 'correspondence theory', in that the typical basis for correspondence relations is grammatical outputs, forms being considered by the evaluation mechanism of the grammar.

[3] Relevant here too is what Hock and Joseph (1996: 293), drawing on the fine work of Samuels 1972, call 'phonesthematic attraction' to describe cases where sound symbolic elements attract other forms into taking on some aspect of their shape (as with early Modern English *sacke* "sink, droop" turning into *sag* through the influence of other words in [-æg] with meanings pertaining to "slow, tiring, tedious action"); since sound symbols can potentially be considered morphemic in nature, the influence in such cases is not just phonic but involves some semantic basis as well.

that they all share the phoneme sequence -VOWEL–*vis* (of which the *–vi-* can be considered a shared stem-forming morpheme).

2 Case Studies in Phonically Based Analogy

The examples presented here range over changes in pronunciation (2.1-2.5),[4] changes in meaning (2.6-2.8), including an example from language contact/bilingualism, changes in lexicon and morphology (2.9-2.10), and changes in syntax (2.11-2.12). In many, perhaps most, of these cases, it is not possible to demonstrate conclusively that sound alone is responsible for the change (though 2.5 comes close), but the aggregate effect of so many examples in which sound seems to have been a relevant dimension to the analogical linkage, I would claim, is to show that a phonic basis for analogy is a distinct possibility that cannot simply be dismissed and thus must be taken into consideration whenever analogy is invoked.

All of the forms cited here are ones that I have heard over the past 30 or so years of collecting interesting examples of language change in action. Although I cannot give precise information about the speakers or the circumstances under which the form was uttered, I vouch for the accuracy of my noting of the forms and note that none is based on a unique instance, and some may even represent longer-standing variation that has been maintained.[5] In each case, I present the facts along with my interpretation of a phonic basis for the analogy, offered without an extensive justification at this stage, in hopes of sparking the necessary weighing of alternative interpretations. Also, where the examples provide the basis for some observations of a more general kind about the nature of analogy, some further comments are included.

2.1 Modern English <memento>

A common pronunciation for the word *memento* "a reminder of the past" in modern American English is [momento] with [o] in the first syllable instead of the 'correct', i.e. historically prior and otherwise expected (note the spelling, for instance) mid-vowel [e]. No similar change is observed in the word *pimento* nor, perhaps more importantly since it involves the same morpheme, in *memorial*, suggesting that the change in *memento* cannot be a regular sound change affecting [e] or [I] between labials, for instance. Presumably, the [o] is based on the word *moment*, which is strongly linked phonically with *memento* due to their sharing the onset of *mVm* and to their both having the sequence *–nt-* following later in the word. Admittedly, there is also a weak semantic link via the phrase *of great moment* and the adjective *momentous***,** both of which mean "memorable" to some extent. More interestingly, one effect of the phonic analogy that leads to [momento] is a severing — or at least weakening — of the once-phonetically compatible linkage between *memento* and *memorial* and other derivatives, or to put it in a different way, the

[4] Note that changes in pronunciation are not the same as sound change, as they may have a variety of causes, including nonphonetic ones, and they need not be regular; I take regularity and purely phonetic conditioning to be the hallmarks of sound change in the strict sense, what may be called 'Neogrammarian sound change' or 'sound change proper' (see Joseph 2008 and Anderson, Dawson, and Joseph 2010: 267 on this latter term).

[5] Such is the case with 2.6 (*flaunt* vs. *flout*), as Henning Andersen (personal communication, 5 October 2004) has brought to my attention (and cf. the OED's citation of *flaunt* in the sense of 'flout' from as early as 1923); so also with 2.3 (*nuclear*) and 2.11 (the *as far as* construction), and possibly others.

morphemic linkage with *memorial*, *remember*, etc. was not strong enough to counteract the effect of the phonic linkage with *moment*.

2.2 Modern English <consonantal>

The adjective associated with the noun *consonant* is *consonantal*, meaning "having to do with a consonant", and while it is generally pronounced, as would be expected from the spelling, [kansənæntəl], there are speakers, such as myself, who regularly say instead [kansənɛntəl]. The source of this innovative pronunciation is obscure, to be sure, but it is presumably based on near-rhyme *continental*; there is here some morphological link in that both *consonantal* and *continental* are denominal adjectives in *–al*, but the main connection between the two is sound-based, via shared onset, shared syllable-count, and shared syllable structure. Moreover, as with *grevis* discussed above, even a morphemic link gives a phonic link, here with respect to the final element *–al*.

2.3 Modern American English <nuclear>

One relevant case that has gotten a fair bit of play over the years in the popular press due to its being, it seems, the pronunciation of choice among American presidents, including Dwight D. Eisenhower, Jimmy Carter, and George W. Bush, is the adjective *nuclear* "having to do with a nucleus" pronounced as [nukyulər], for etymologically correct [nukliyər]. Here the influence seems to be the class of adjectives like *popular*, *particular*, *insular*, etc., with *nuclear* in essence 'assimilating' to, i.e. being attracted into, the class of adjectives in *–lar*. However, even if the end-point is a morphological type with a suffixal *–lar*, the starting point has to be the phonic form, with no strong morphemic basis. That is, even though *nucleus* has an *–l-* in it, its *–l-* has a different placement and morphemic status from that seen in *people/populace*, *particle*, etc., i.e. in the base words for *popular*, *particular*, etc.[6] The phonic form that gives a starting point for the attraction is *...lər* in both the attractor and attractee, discontinuous in the case of *nuclear* (thus *...l...ər*), and the end result is *...(yu)lər* in both.

2.4 Modern American English <extraterrestrial>

The adjective *extraterrestrial* "from outer space" is innovatively pronounced by some speakers as ending in [...stiyəl] as opposed to the etymologically correct ending [...striyəl]); the basis here seems to be attraction to, that is to say influence from, *celestial* 'heavenly', with the phonic link being the shared sounds [–ɛst…iyəl], though admittedly there is a semantic connection as well between these words.

2.5 Modern American English <academia>

One particularly intriguing case is the pronunciation of *academia* as [ækəd**ej**miə] (at least in American English) as opposed to the more usual [ækəd**ij**miə]. In talking about this case over the years, in classes or in presentations, I have been told that it is a pseudo-

[6] Dr. Tom Stewart (personal communication, Spring 2001) has told me that the noun *nucleus* can be heard as [nukjuləs], and I have personally verified that since. Though this could be the basis for the adjectival pronunciation discussed here, I am inclined to think — since there is no obvious (to me) basis for [nukjuləs] in and of itself — that the noun here is a back-formation derived from the innovative pronunciation of the adjective.

learnedism, affecting a Latin-like style of pronunciation or an Italian- or a Spanish-like one, but that ignores the basic point of why this word out of other possible words would have been affected, and why that particular affectation as opposed to other possible alterations occurred with this word. That is, there are other learnèd words that do not undergo a similar fate, such as *anemia*, for which there is no variant [ən**ej**miə],[7] or even *epidemiology*, with the same surrounding environment as *academia* (i.e., with *d* and *m* flanking the affected vowel), for which there is no [ɛpɪd**ej**mi…].[8] Nor can influence from a morphologically or semantically related word be responsible; in fact, one does hear on occasion [ækəd**ɛ**miə], based on the pronunciation of *academic*, but there is no obviously related word with [-ej-]. But when one looks to less obvious (but, I would argue, no less relevant) forms, a solution awaits; thus, I suggest that this innovative pronunciation of *academia* is based on the influence of *macadamia (nut)* where the basis for the connection is purely phonic in nature – the relevant phonic links are the large number of shared segments in the same order, in particular, [ækəd…miə]) and the shared rhythmic stress pattern. This influence seems to be felt even though there is no semantic connection whatsoever; phonic form alone seems to matter here.[9]

2.6 American English <flaunt>

The verb *flaunt*, canonically having the meaning "show off; display ostentatiously", can now be used as well quite commonly (though prescriptively 'incorrectly') in meaning of "show contempt for; scorn". This innovative meaning is exactly the meaning of *flout*, which, not coincidentally I would argue, is phonically similar to *flaunt* in that both share [fl…t]. Thus, with this innovative meaning, *flaunt* has assimilated in meaning to *flout*, where the link between the two, the basis for the analogical influence of *flout* over *flaunt*, is a shared aspect of their phonic shape.

2.7 American English <diffident>

Somewhat similar to *flaunt* is the situation with *diffident* 'shy, lacking in self-confidence", in that it is now used by some speakers in the meaning of "having no interest in or concern for". Presumably what has happened here is that *diffident* has been 'attracted' by the phonically similar *indifferent*, which has that very meaning; crucial here to the attraction is the fact that the two words share the syllable [...dɪf...]), which is

---

[7] My good friend and many-time collaborator Richard Janda and I independently came up with this idea about the source of the innovative pronunciation of *academia*, at some point in the mid-to-late 1980s, and we have each since used it in classes and in presentations. My including it here in print is with Rich's permission, and in fact, I must acknowledge his input through enlightening discussion we have had on this example, including the particular point about *anemia*; I have benefited greatly too from the many general discussions Rich and I have had over the years concerning not just *academia* but the whole overall line of reasoning adopted herein as well.

[8] Henning Andersen (personal communication, 5 October 2004) tells me that the word *schizophrenia*, widely pronounced as ending in […ijniə], can be heard also as ending in […ejniə], suggesting that there may indeed be a 'learnèd word' pronunciation coming to be associated with [ej] in certain items. I am inclined however to think of possible influence from semantically (and somewhat phonically) connected *mania* in this case, though the nature of this sort of variation in general is such that one cannot rule out any of the possible pressures.

[9] One is inevitably led to make a quip about academics being nuts, and indeed, I even own a T-shirt, a gift from a former student, Dr. Halyna Sydorenko of Toronto, that says "Academia Nut". Such a connection seems unlikely to have played a role here, however. (To see a photo of me in the shirt, go to http://osu.academia.edu/BrianDJoseph.)

stressed in each, as well as having the same end segments and, except for the prefix *in-*, the same rhythmic structure.

2.8 American Norwegian <brand>

The same effect as that seen in 2.6 and 2.7 can be observed in language contact, where the 'attraction' takes place across languages whose speakers are in contact and are bilingual. In particular, Haugen 1969 has noted what he calls 'homophonous extensions', exemplified by American Norwegian *brand*, which has the meaning "bran", as opposed to the meaning "fire" in Standard Norwegian, due, in his account, to the influence of American English *bran*. Haugen's use of the descriptor 'homophonous' signals his recognition of the relevance of the phonic link between the attractor and the attractee.

2.9 American English <as of yet>

Although meaning can be affected by phonically based analogical attraction, as the examples in 2.6-2.8 show, the results of such analogical pressure need not always make sense. Rather, it can effect changes in the form alone even if aspects of the meaning are altered in unusual ways. A case in point is the expression *as of yet*, which seems to be an innovative crossing, a contamination that is, between two phrases, *as yet* and *as of now*, that were already present in the language. The emergence of *as of yet* means that either *as yet* has taken on *of* due to influence from *as of now*, or else *as of* has taken on *yet* as a possible complement due to influence from *as yet*. In either case, there is a shared phonic link through the word *as*, but there is as well a semantic link in that both are time expressions. Nonetheless, along with the analogical assimilation that leads one of these expressions in the direction of the other with regard to form, there is either a complication of or a shift in the semantics of the relevant pieces.

In particular, in the phrase *as of X*, the complement *X* generally has a definite and fixed time reference of some sort (e.g. *as of December*, *as of now*, *as of 3:33PM*, etc.); however, in the innovative *as of yet*, the complement has a very different kind of time reference, certainly not anything that could be characterized as definite in any sense, and thus a complication. Alternatively, one could say that the meaning of *yet* has shifted to accommodate its use in a new expression or that the requirements of *as of* have changed so as to allow a referentially vague term like *yet* as a complement. Either way there is a change beyond the new form, and the analogy leading to *as of yet*, with its phonic basis, is in large part responsible.

2.10 Latin <queō, nequeō>

The Latin verbs *queō* "I am able" and *nequeō* "I am not able" may well show morphological developments that under one account of their etymology would be a case of phonically based analogy. The standard etymology[10] treats *nequeō* as the older form, deriving from *neque* "and not" plus *eō* "go", originally in an impersonal passive formation *nequitur* "it does not go", with *queō* then a back formation created by slicing off the clear negative morpheme *ne-*. This suggestion fits the facts formally and may well be right, but it is not necessarily the most satisfying possibility on the semantic side. As an alternative, one might look to a different root as underlying these verbs, in particular

[10] See Ernout & Meillet (1939: s.v.).

Proto-Indo-European *k^{w}ey- "make, do" (as seen in Greek *poiéō*), so that the sense "be unable" would stem from "not to be done" (that is, "not doable"). While admittedly speculative (as is often the case with root etymologizing), in that case, these verbs would show 'assimilation' in their inflection to the form of *eō* "I go", in the following ways: from a preform 1PL.PRES *k^{w}ey-o-mos, one would expect either Latin *queumus** (with the phonetic development of *-eyo- seen in *ey-ont- "going" => *eunt-*) or *quēmus** (with the analogical development seen in forms like *monēmus* "we warn"). Instead what occurs for these verbs is *(ne)quīmus*, with the same root form as *īmus* "we go", from *ey-mos. Similarly, the infinitive is *(ne)quīre*, just like *īre* "to go", even though the expected outcome would be something like *(ne)quēre**. This cannot be proven conclusively, and it may well be that Ernout and Meillet are right in linking these verbs etymologically with *eō* from the start, but if the semantic connection is considered suspect, so that an alternative etymology is sought, then the later issue of how *(ne)queō* came to be linked with *eō* would have to be based not on their semantics but on the fact that they rhyme. That is, what would link the verbs, in this interpretation, and make the analogical influence possible, therefore, would be a phonic connection.

2.11 English <as far as…>

An example involving phonic analogy that affects syntax can be seen in the changes discussed by Rickford et al. 1995 with regard to the English construction beginning with *as far as* and signalling a focalized element. In particular, they note that a clear old construction in Modern English is that illustrated in (1), in which following the focalized element preceded by *as far as*, there is a verbal coda, usually *be concerned* though others such as *go* can also be found:

(1) a. As far as John is concerned, forget about him!
b. As far as John goes, forget about him!

In addition to this construction, there is another one, which Rickford et al. quite appropriately take to be innovative, in which *as far as* occurs but the verbal coda is lacking, as in (2):

(2) As far as John, forget about him!

Their main concern is the spread of the innovative construction in the past 200 years and especially in the later half of the 20th century, but they discuss various possible explanations for the appearance of the innovative pattern in the first place. One that they consider to be possible, but which in my view they pass over a bit too hastily (p. 115), is that given by Faris 1962 concerning possible involvement of another focalizing construction with *as for*, as in (3), which has no verbal coda:

(3) a. As for John, forget about him!
b. *As for John is concerned, forget about him!

The absence of the verbal coda in the *as for* construction would provide a model for its analogical absence in the *as far as* construction. But what is the basis for a connection between the two constructions? They are functionally linked, of course, in that both mark focused elements, but alongside this functional connection, there is another that cannot be ruled out, namely what Faris may have been hinting at when he referred to the influence

of 'the closely resembling *as for'* (p. 238): a phonic link. That is, one could claim that *as for* provided a suitable model for *as far as* based on the shared phonic form between the two of *as* and *f-V-r*. In this way, the innovative verbless construction would be a contamination or crossing (as seen above), an analogical creation with a phonic basis.

2.12 American English <being that>

My final example also is a case of syntax being affected by a phonically based analogy, and is quite parallel to the *as far as* example in 2.11. In this instance, the two older constructions that play a role, in my account, are the subordinate clauses (underlined) exemplified in (4):

(4) a. Seeing that John is here, we can start.
b. It being the case that John is here, we can start.

and the innovative construction is that illustrated in (5):

(5) Being that John is here, we can start.

All of these represent ways of stating the circumstances under which the action of the main clause occurs, (4a) with a gerund (or participle) that ostensibly is linked to the main clause subject and (4b) with an absolute construction containing an expletive *it* serving as subject to *being*. In the case of (5), there is as well a 'dangling' participle, in that *being* is not linked to any main clause argument, but also the syntactic anomaly of the suppression of the expletive subject of *being*, even though English in general is not a *pro*-Drop language.[11] How did the innovative construction in (5) arise? It is my contention that it is the result of a crossing of the two older constructions seen in (4), where the connection between the two is on the one hand functionally based in that both indicate attendant circumstances, but further, that it is aided by the phonic link between the two as well, essentially the rhyming of *seeing* with *being*, and thus due to the same sort of pressures that gave rise to the innovative *as far as* construction, and indeed the other innovative forms throughout section 2.

3 Conclusion

There is more that can be said about these examples and their collective effect. For instance, in some cases, the analogy results in a new form that is far from regular or simplified, far from 'optimal', as with the *being that* construction in 2.12, with its odd suppression of a pronoun that runs counter to otherwise quite general English subject requirements, or *as of yet* with its odd semantics or selectional anomaly. The suggestion that these anomalies emerge by analogical pressures means that analogy cannot be taken as an optimizing or regularizing force per se, except perhaps when applied to individual cases; that is, rather than leading to system-wide regularization and simplification ("system optimization" in the sense of Kiparsky 2000), analogy can introduce

[11] Admittedly, most treatments of *pro*-Drop refer to the suppression of subject pronouns in finite clauses (as in Modern Greek *tréxo* "I am-running" (literally, "am-running"). However, English has the free suppression of subject pronouns only in imperatives, and gerund/participial forms normally lack a subject only under circumstances of control from a main clause nominal (as in (4a)). Thus the absence of *it* here is innovative from a syntactic point of view.

complication into the system – the regularization would seem to be just on a very localized basis (in the sense of Joseph & Janda 1988), in that, as here, there is an 'inner logic', as it were, to the creation of *as of yet* because of the presence of *as yet* and *as of now*, at least in terms of its surface form; so also with *being that*.[12]

Finally, it must be emphasized that even though phonic effects, based on these examples, seem to be capable of playing an important role in establishing analogical links among forms, it is not the case that phonic effects hold sway every time one of these forms is uttered. Rather, as with any change, once a new form takes hold, the path by which it arrived at that particular shape is largely irrelevant. For instance, even though the currently widespread American English pronunciation of *often* with medial *–t-* has its origins in a spelling-based pronunciation, it is not the case that every time it is uttered now, speakers have the spelled form in mind inducing them into pronouncing the *–t-*; rather for most such speakers, *often* is simply learned with a *t* and thus always pronounced that way. So too in the examples discussed here: it is not the case that every innovative utterance of *academia* has a *macadamia* lurking behind it, so to speak. However, as the need to separate the impetus of an innovation from its spread is necessary in most accounts of language change in general, this aspect of the discussion merely places these examples in conformity with what is known about language change more generally.

References

Anderson, John; Hope C. Dawson; and Brian D. Joseph. 2010. Historical linguistics. *The Routledge linguistics encyclopedia* (3rd edition), ed. by Kirsten Malmkjaer, 225-251. London: Routledge Publishers.

Anttila, Raimo. 1972/1989. *An introduction to historical and comparative linguistics*. New York: Macmillan (2nd edn., *Historical and comparative linguistics*, Amsterdam: John Benjamins,1989).

Anttila, Raimo. 1977. *Analogy* (Trends in linguistics. State-of-the-art reports). The Hague: Mouton.

Anttila, Raimo. 2003. Analogy: The warp and woof of cognition. In Joseph and Janda, 425-440.

Anttila, Raimo and Warren Brewer. 1977. *Analogy: A basic bibliography***.** Amsterdam: John Benjamins Publishing Co.

Dawson, Hope C. 2005. *Morphological variation and change in the Rigveda: The case of* –au *vs.* –ā. Columbus: The Ohio State University Ph.D. Dissertation.

Ernout, Alfred and Antoine Meillet. 1939**.** *Dictionnaire étymologique de la langue latine. Histoire des mots*. Paris: Klincksieck**.**

Faris, Paul. 1962. 'As far as halfbacks, we're all right'. *American Speech* 37.236-38.

Haugen, Einar. 1969. *The Norwegian language in America. A study in bilingual behavior*. Bloomington: Indiana University Press.

Hock, Hans Henrich. 2003. Analogical change. In Joseph and Janda, 441-460.

Hock, Hans Henrich, and Brian D. Joseph. 1996. *Language history, language change, and language relationship. An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.

[12] This line of reasoning is pursued further in Joseph 2012.

Joseph, Brian D. 2008. Historical Linguistics in 2008: The state of the art. *Unity and Diversity of Languages*, ed. by Piet van Sterkenberg, 175-187. Amsterdam: John Benjamins Publishers.

Joseph, Brian D. 2012. Optimality, optimization, and analogy. Columbus: Ohio State University, ms.

Joseph, Brian D., and Richard D. Janda. 1988. The how and why of diachronic morphologization and demorphologization. *Theoretical morphology: Approaches in modern linguistics*, ed. by Michael Hammond and Michael Noonan, 193-210. San Diego: Academic Press.

Joseph, Brian D., and Richard D. Janda, eds. 2003. *The handbook of historical linguistics*. Oxford: Blackwell Publishers.

Kiparsky, Paul. 2000. Analogy as optimization: 'Exceptions' to Sievers' Law in Gothic. *Analogy, levelling, markedness. Principles of change in phonology and morphology*, ed. by Aditi Lahiri, 15-46. Berlin: Mouton de Gruyter.

Miller, D. Gary. 1982. *Homer and the Ionian epic tradition: Some phonic and phonological evidence against an Aeolic phase*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Rickford, John R.; Thomas A. Wasow; Norma Mendoza-Denton; and Juli Espinoza. 1995. Syntactic variation and change in progress: Loss of the verbal coda in topic-restricting *as far as* constructions. *Language* 71.1.102-31.

Samuels, M. L. 1972. *Linguistic evolution; With special reference to English*. Cambridge: Cambridge University Press.

Vennemann, Theo. 1972. Phonetic analogy and conceptual analogy. *Schuchardt, the Neogrammarians, and the transformational theory of phonological change: Four essays by Hugo Schuchardt, Theo Vennemann, Terence H. Wilbur* (Linguistische Forschungen, 26), ed. by Theo Vennemann and Terence H. Wilbur, 181-204. Frankfurt am Main: Athenäum.

Watkins, C. 1995. *How to kill a dragon. Aspects of Indo-European poetics*. New York: Oxford University Press.

[joseph.1@osu.edu]

COORDINATION IN HYBRID TYPE-LOGICAL CATEGORIAL GRAMMAR

Yusuke Kubota and Robert Levine
University of Tokyo and Ohio State University

Abstract

We formulate explicit analyses of certain non-standard coordination examples discussed in Levine (2011) in a variant of categorial grammar called *Hybrid Type-Logical Categorial Grammar* (Kubota 2010; Kubota & Levine 2012; Kubota to appear). These examples are of theoretical importance since they pose significant challenges to the currently most explicit and most comprehensive analysis of coordination, formulated in a variant of HPSG called Linearization-based HPSG (Reape 1996; Kathol 1995) and advocated by various authors in the recent literature (Yatabe 2001; Crysmann 2003; Beavers & Sag 2004; Chaves 2007; Sag & Chaves 2008). This approach, which we call the Linearization-Based Ellipsis (LBE) approach to coordination, builds on the key idea that apparent non-standard coordination all reduce to constituent coordination under surface ellipsis. The seemingly heterogeneous set of data catalogued in Levine (2011), involving different types of non-standard coordination, uniformly point to an analysis in which the apparently incomplete constituents that are coordinated in the overt string are in fact complete (i.e. non-elliptical) constituents with full-fledged semantic interpretation, thus directly counterexemplifying the predictions of ellipsis-based approaches including the LBE variant. The sophisticated syntax-semantics interface of the framework

Yusuke Kubota and Robert Levine. Coordination in Hybrid TLCG. *OSUWPL* Volume 60, Spring 2013, pp. 21-50.

we propose in this paper straightforwardly captures the interactions between such non-standard coordination and various scopal expressions, demonstrating the real empirical payoff of the direct coordination analysis of non-standard coordination (of the kind widely adopted in categorial grammar) that has not been fully recognized in the previous literature.

1 Introduction

Levine (2011) provides a thorough critique of an approach to coordination that has become prominent in the recent literature of HPSG. The key idea of this approach, which we call the Linearization-Based Ellipsis (LBE) approach to coordination (Yatabe 2001; Crysmann 2003; Beavers & Sag 2004; Chaves 2007; Sag & Chaves 2008), is to analyze a wide range of coordination examples that apparently pose problems for phrase structure-based theories of syntax such as HPSG via a single mechanism of surface ellipsis. The analysis is technically implemented in a variant of HPSG that relaxes the mapping between the combinatoric structure and surface string known as Linearization-based HPSG (Reape 1996; Kathol 1995). In this paper, we take up some of the key examples from Levine (2011), which pose serious problems for the LBE approach and provide explicit analyses of them within a variant of categorial grammar called *Hybrid Type-Logical Categorial Grammar*. (Kubota 2010; Kubota & Levine 2012; Kubota to appear). While building heavily on ideas from previous literature of categorial grammar (CG), the proposed framework is novel in that it recognizes *both* the directionality-sensitive mode of implication (i.e. forward and backward slashes) familiar from the tradition of Lambek (1958) and the directionality-insensitive mode of implication employed in some of the more recent variants of CG pioneered by Oehrle (1994). This hybrid implication architecture enables a flexible and sophisticated syntax-semantics interface that is not available in previous variants of CG, and we show below that the analyses of the empirical phenomena taken up in this paper crucially exploits this flexibility and systematicity of the proposed framework. As will become clear below, due to the generality of the underlying logic, the present framework is not amenable to the kinds of criticisms that have been occasionally raised against previous variants of CG (especially CCG), which, for various reasons, do not entertain the flexibility of the logic-based syntax-semantics interface characteristic to CG in a fully general way.

The key hypothesis of the LBE approach to coordination, in its strongest version, is that a wide range of non-standard coordination such as the following (dependent cluster coordination (DCC) in (1a), right-node raising (RNR) in (1b), agreement anomaly in nominal head coordination in (1c) and unlike category coordination (UCC) in (1d)) can all be reduced to ordinary constituent coordination via surface ellipsis along the lines of (2).

(1) a. I gave Robin a book and Terry a pair of pliars.
b. I gave Robin, and Leslie offered Terry, a pair of pliars.
c. That man and woman are arguing again.
d. Robin is a Republican and proud of it.

(2) a. [$_S$ [$_S$ I gave Robin a book on Thursday] and [$_S$ ~~I gave~~ Leslie a book on Friday.]]
b. [$_S$ [$_S$ I gave Robin ~~a pair of pliars~~] and [$_S$ Leslie offered Terry, a pair of pliars]].
c. [$_{NP}$ [$_{NP}$ That man] and [$_{NP}$ ~~that~~ woman]] are arguing again.
d. [$_S$ [$_S$ Robin is a Republican] and [$_S$ ~~Robin is~~ proud of it.]]

The strikeout in (2) is meant to represent purely phonological deletion which is licensed on the condition that the same string appears in the other conjunct. This deletion operator (at least on the null hypothesis) is supposed to affect only the pronunciation of the sentence and not its semantic interpretation.

It should be noted that not all advocates of the LBE approach endorse an elliptical analysis for all of these cases (see, for example, Yatabe (2012), who expresses the view that an ellipsis-based approach is not appropriate for the latter two cases). However, since the plausibility of the hypothesis in part depends on its generality and since ellipsis-based analyses along the lines of (2) have been suggested for all of these phenomena by at least some authors advocating the LBE approach, we include them all here for completeness.

Following Levine (2011), we point out below that *none* of these cases are amenable to ellipsis-based analyses once we extend the dataset to cases in which the semantics of the coordinated expressions have non-trivial consequences for the compositional semantics of the whole sentence. The relevant examples involve the interactions between coordination and scopal expressions that appear outside the coordinate structure. We show below that the predictions of the LBE approach is systematically falsified in each such case. We then present explicit compositional analyses of these phenomena in Hybrid TLCG. We show that independently motivated analyses of each of these constructions interact properly with analyses of scopal elements in the proposed framework to yield the correct predictions in the relevant examples straightforwardly. Besides providing a general argument for a CG-based analysis of coordination, the data discussed in this paper thus provides strong empirical evidence for the proposed variant of categorial grammar (among other variants) in that the interactions between coordination and scopal expressions that they manifest call for exactly the sort of hybrid architecture of the syntax-semantics interface that is unique to it.

2 Contraindications for Linearization-Based Ellipsis (LBE)

2.1 Symmetrical, respective and summative predicates and nonconstituent coordination

Perhaps the strongest piece of evidence against ellipsis-based approaches to coordination comes from data such as (3a) and (4a), long known in the literature as a paradigm case which demonstrates the inadequacies of ellipsis-based analyses of nonconstituent coordination such as DCC and RNR (Abbott 1976; Jackendoff 1977; Gazdar 1981).

(3) a. I said $\left\{\begin{array}{l}\text{the same thing}\\ \text{different things}\end{array}\right\}$ to Robin on Thursday and (to) Leslie on Friday.

b. I said $\left\{\begin{array}{l}\text{the same thing}\\ \text{different things}\end{array}\right\}$ to Robin on Thursday and <u>(I) said $\left\{\begin{array}{l}\text{the same thing}\\ \text{different things}\end{array}\right\}$ to Leslie on Friday.</u>

(4) a. Robin reviewed, and Leslie read, $\left\{\begin{array}{l}\text{the same book}\\ \text{different books}\end{array}\right\}$.

b. Robin reviewed <u>$\left\{\begin{array}{l}\text{the same book}\\ \text{different books}\end{array}\right\}$</u>, and Leslie read $\left\{\begin{array}{l}\text{the same book}\\ \text{different books}\end{array}\right\}$.

Ellipsis-based analyses predict that the NCC examples in (3a) and (4b) are synonymous to the their constituent coordination counterparts in (3b) and (4b) (from which the NCC examples are derived by deleting the underlined parts—underlines in examples mean the same thing in all of the examples below), but this prediction is not borne out. (3a) and (4b) exhibit the so-called *internal reading* of *same* and *different* (Carlson 1987), which simply asserts the (non-)identity of the thing(s) in question. The constituent coordination counterparts in (3b) and (4b) have only the anaphoric, *external reading*, which presupposes the existence of some entity already salient in the discourse and asserts the (non-)identity between the things in question and that discourse-salient entity.

This type of non-parallel between NCC and their alleged clausal counterparts is not limited to symmetrical predicates, but is actually much more widespread. As noted, for example, by Abbott (1976) and Chaves (2012), essentially the same pattern is observed with the so-called 'respect readings' of sentences involving the adverb *respectively*, and the summative interpretations of numerical expressions such as *a total of $1000*. Examples of DCC are given in (5) and (6). Similar examples of RNR can be constructed easily.

(5) a. I lent *Barriers* and *Syntactic Structures* to Robin on Thursday and (to) Leslie on Friday, respectively.

b. I lent *Barriers* and *Syntactic Structures* to Robin on Thursday and <u>I lent *Barriers* and *Syntactic Structures* (to) Leslie on Friday, respectively.</u>

(6) a. I lent *$1000 in total* to Robin on Thursday and (to) Leslie on Friday.

b. I lent *$1000 in total* to Robin on Thursday and <u>I lent *$1000 in total*</u> (to) Leslie on Friday.

Here, too, the NCC examples have meanings that are not available in their alleged clausal counterparts, and this fact remains a mystery for ellipsis-based approaches to coordination.

It turns out that working out an explicit compositional semantics for this type of examples poses a significant challenge for any type of approach to NCC, be it ellipsis-based or not. In fact, as far as we are aware, except for Kubota (2010), there is no explicit proposal in the literature that provides a completely satisfactory solution for this problem. We

reproduce in section 3 Kubota's (2010) analysis of examples like (3a) and (4a) (involving symmetrical predicates), which exploits the hybrid implication architecture of the present framework.

2.2 UCC and extraction

The use of ellipsis to derive examples like (7a) from underlying 'source' structures like (7b) and thereby eliminate the 'unlikeness' of UCC was suggested by Beavers & Sag (2004) and then advocated more extensively by Chaves (2006).

(7) a. Robin is a Republican and proud of it.
b. Robin is a Republican and (Robin) is proud of it.

However, such an analysis leads to severe mispredictions once one considers more complex examples. Note first that strings like *rich and a Republican*, which exemplify the unlike category coordination, can be topicalized as in (8a):

(8) a. Rich and a Republican, Robin definitely is t.
b. Rich Robin definitely is t and a Republican Robin definitely is t.

The only way to derive (8a) from ellipsis is to assume an underlying source of the form in (8b) involving conjunction of full-fledged clauses. On this type of analysis, the following examples turn out to be crucially problematic:

(9) a. (Both) poor and a Republican, you can't possibly be t.
b. (Both) poor you can't possibly be t and a Republican you can't possibly be t.
(10) a. Dead drunk but in complete control of the situation, no one can be t.
b. Dead drunk no one can be t but in total control of the situation, no one can be t.

The ellipsis-based approach demands (9a) and (10a) to be derived from (9b) and (10b), but there is a mismatch in semantic interpretation that is essentially parallel to the symmetrical predicate data from the previous section. In (9a) and (10a), the modal scopes over the whole coordinated string *rich and a Republican*; in other words, in (9a), what is negated is the property of simultaneously being poor and a Republican. The non-elliptical sources in (9b) and (10a) lack that interpretation totally; they can only be interpreted as a conjunction of negation, which has a stricter truth conditions than their alleged elided counterparts.

The same observation can be replicated in yet another displacement construction, namely, pseudocleft. In (11a), a conjoined unlike category occupies the focus position of a

pseudocleft sentence. On the ellipsis-based analysis, this example has to be derived from the underlying source in (11b), but again, there is a semantic mismatch between the UCC example and its alleged underlying source.

(11) a. What you cannot become (simultaneously) is highly intelligent and (yet) a raving fundamentalist.
b. What you cannot (simultaneously) become is highly intelligent and (yet) what you cannot (simultaneously) become is a raving fundamentalist.

2.3 Nominal head coordination under a singular determiner

Finally, we consider the case of nominal head coordination under a singular determiner exemplified by data such as (12):

(12) That man and woman are arguing again.

Chaves (2007) and Sag & Chaves (2008) suggest the possibility of deriving (12) from an underlying source of the form of (13). An apparent advantage of such an analysis is that it provides an immediate (and simple) solution for the seemingly anomalous agreement pattern in (12) (where a singular determiner is used for an NP which clearly refers to multiple individuals).

(13) That man and that woman are arguing again.

Consideration of a wider range of data, however, once again reveals that such an ellipsis-based analysis is too simplistic. Examples like the following noted by Heycock & Zamparelli (2005) in which a symmetrical modifier appears in the the 'ellipsis' environment resist an analysis along the lines of (13).

(14) a. That ill-matched man and woman are fighting again.
b. *That ill matched man and that ill-matched woman are fighting again.

(15) a. That mutually hostile judge and defense attorney were constantly sniping at each other during the trial.
b. *That mutually hostile judge and that mutually hostile defense attorney were constantly sniping at each other during the trial.

The alleged underlying sources for (14a) and (15a), given in (14b) and (15b), are simply ungrammatical. Here again, the problem essentially stems from the fact that an ellipsis-based analysis gets the semantic scope wrong. For example, to derive the right meaning for

(14a), the symmetrical predicate *ill-matched* has to scope over the conjoined noun *man and woman* so as to establish the right relation between the man and the woman in question. Such an interpretation cannot be obtained from the 'source' structure (14b), where two distinct tokens of *ill-matched* modify the nouns *man* and *woman* separately within each conjunct.

2.4 Summary

The LBE approach to coordination at first sight appears to provide a simple and uniform solution for a wide range of apparently heterogeneous set of non-standard coordination phenomena. However, as we have seen above, the success is illusory. Once we look beyond the simplest cases exemplified by (1), the ellipsis-based approach faces several severe difficulties. Specifically, in all of the cases discussed in this section, whose key examples are repeated in (the a.-examples of) (16)–(18), we see a systematic interaction between the coordinated expression and scopal operators that appear *outside* the coordinate structure in the overt string.

(16) a. I said { the same thing / different things } to Robin on Thursday and (to) Leslie on Friday.

b. I said { the same thing / different things } to Robin on Thursday and ~~(I) said~~ { **~~the same thing~~** / **~~different things~~** } ~~to~~ Leslie on Friday.

(17) a. (Both) poor and a Republican, you can't possibly be t.

b. (Both) poor ~~you **can't possibly** be~~ t and a Republican you can't possibly be t.

(18) a. That ill-matched man and woman are fighting again.

b. That ill-matched man and ~~that **ill-matched**~~ woman are fighting again.

In all these cases, the observed empirical pattern is that the operator scopes over the whole coordinate structure in a way that mirrors the surface form of the sentence. The LBE approach systematically mispredicts in such cases, since, on the ellipsis-based analysis, the scopal operator is part of the elided string (as in b. above) and hence appears *inside* each conjunct, in effect reversing the scopal relation between the operator and the coordinate structure from what is actually observed.

3 Hybrid Type-Logical Categorial Grammar

The central characteristic of the variant of CG that we propose in this paper is that it recognizes two kinds of implication, namely, the order-sensitive forward and backward slashes familiar from the tradition of Type-Logical Categorial Grammar originating from Lambek's (1958) work, and the order-insensitive mode of implication tied to phonological λ-binding in more recent variants of CG (stemming from Oehrle's (1994) work) that

relegate word order-related information from the combinatoric component of syntax to a separate morpho-phonological component. As will become clear below, the hybrid architecture of the present framework is exploited crucially in capturing the interactions between coordination and various scopal expressions. Directional variants of categorial grammar provide an elegant analysis of non-standard coordination (especially nonconstituent coordination including both DCC and RNR), while they are subotimal for scopal phenomena due to the inherently directional nature of the underlying calculus. By contrast, variants of CG that relegate word order entirely to a separate prosodic component enables a straightforward treatment of scopal phenomena, but they have the drawback that the elegant analysis of (non-standard) coordination in directional variants of CG is lost, due to the fact that syntactic categories of linguistic expressions do not carry order-related information (an aspect of directional variants of CG that is crucially exploited in the analysis of coordination). Hybrid TLCG entertains the advantages of both directional and non-directional variants of CG, by recognizing both kinds of implication within a single calculus. The complex interactions between coordination and scopal expressions exhibited by the data observed in section 2 requires exactly this kind of architecture, providing empirical evidence for the novel architecture of CG embodied in the proposed framework.

3.1 Hybrid Implication System as an Underlying Logic

Following Oehrle (1994), we write linguistic expressions as tuples of phonological form, semantic interpretation and syntactic category (written in that order). Our system recognizes both directional modes of implication ($/$ and $\backslash$) and a non-directional mode of implication that we call the *vertical slash* ($|$, for which we write the argument to its right, just as with $/$). The full set of inference rules posited in the calculus, consisting of the Introduction and Elimination rules for the three kinds of slashes, are given in (19).

(19)

Connective	Introduction	Elimination
$/$	$\dfrac{\begin{array}{c}\vdots\ [\varphi; x; A]^n\ \vdots \\ \vdots\quad\ \vdots\quad\ \vdots\end{array}}{\dfrac{b \circ \varphi; \boldsymbol{f}; B}{b; \lambda x.\boldsymbol{f}; B/A}\,/\mathrm{I}^n}$	$\dfrac{a; \boldsymbol{f}; A/B \quad b; \boldsymbol{g}; B}{a \circ b; \boldsymbol{f}(\boldsymbol{g}); A}\,/\mathrm{E}$
$\backslash$	$\dfrac{\begin{array}{c}\vdots\ [\varphi; x; A]^n\ \vdots \\ \vdots\quad\ \vdots\quad\ \vdots\end{array}}{\dfrac{\varphi \circ b; \boldsymbol{f}; B}{b; \lambda x.\boldsymbol{f}; A\backslash B}\,\backslash\mathrm{I}^n}$	$\dfrac{b; \boldsymbol{g}; B \quad a; \boldsymbol{f}; B\backslash A}{b \circ a; \boldsymbol{f}(\boldsymbol{g}); A}\,\backslash\mathrm{E}$
$\lvert$	$\dfrac{\begin{array}{c}\vdots\ [\varphi; x; A]^n\ \vdots \\ \vdots\quad\ \vdots\quad\ \vdots\end{array}}{\dfrac{b; \boldsymbol{f}; B}{\lambda\varphi.b; \lambda x.\boldsymbol{f}; B\lvert A}\,\lvert\mathrm{I}^n}$	$\dfrac{a; \boldsymbol{f}; A\lvert B \quad b; \boldsymbol{g}; B}{a(b); \boldsymbol{f}(\boldsymbol{g}); A}\,\lvert\mathrm{E}$

The key difference between the directional slashes ($/$ and $\backslash$) and the non-directional slash ($|$) is that while the Introduction and Elimination rules for the former refer to the phonological forms of the input and output strings (so that, for example, the applicability of the /I rule is conditioned on the presence of the phonology of the hypothesis φ on the right periphery of the phonology of the input $b \circ \varphi$),[1] the rules for the latter are not constrained that way. For reasoning involving $|$, the phonological terms themselves fully specify the ways in which the output phonology is constructed from the input phonologies. Specifically, for $|$, the phonological operations associated with the Introduction and Elimination rules mirror exactly the semantic operations for these rules: function application and λ-abstraction, respectively. We assume that the binary connective $\circ$ in the phonological term calculus represents the string concatenation operation and that $\circ$ is associative in both directions. For notational convenience, we implicitly assume the associativity axiom $(\varphi_1 \circ \varphi_2) \circ \varphi_3 \equiv \varphi_1 \circ (\varphi_2 \circ \varphi_3)$ and leave out all the brackets indicating the internal constituency of complex phonological terms.[2] The phonological term calculus is a lambda calculus, and we also take the equivalence between β-reduced and unreduced terms to be an axiom.[3]

It should be clear from the way the above rules are formulated that the present system without the rules for $|$ is equivalent to the Lambek calculus (Lambek 1958), while the system with only the rules for $|$ is essentially equivalent to the term-labelled calculus of Oehrle (1994), Lambda Grammar (Muskens 2003), Abstract Categorial Grammar (de Groote 2001), and Linear Grammar (Pollard 2011), with some irrelevant implementational details aside.

3.2 Basic Analyses of Scope and Coordination

As was first demonstrated by Oehrle (1994), λ-binding in the phonological component enables an insightful analysis of quantifier scope, a problem whose general solution has turned out to pose significant theoretical challenges to directional variants of

[1] In this respect, the present calculus follows most closely Morrill & Solias (1993) and Morrill (1994); see Moortgat (1997) and Bernardi (2002) for an alternative formulation where sensitivity to directionality is mediated through a presumed correspondence between surface string and the form of structured antecedents in the sequent-style notation of natural deduction.

[2] For a more fine-grained control of surface morpho-phonological constituency, see Kubota & Pollard (2010) (and also Muskens (2007) for a related approach), which formalizes the notion of multi-modality from the earlier TLCG literature (Moortgat & Oehrle 1994; Morrill 1994) by modelling the mapping from syntax to phonology by means of an interpretation of (phonological) λ-terms into preorders.

[3] Note that the equivalence enforced by the associativity axiom is a property of the prosodic calculus directly reflecting the structures of the objects that the prosodic terms are supposed to model whereas equivalence under β-reduction is a formal property of the calculus itself. In this sense, the two equivalence relations we implicitly assume here are of somewhat different nature. In particular, the latter assumption underlies the fundamentally hybrid nature of the proposed system of tripartite inference as a whole in that β-reductions of prosodic terms that result from inferences involving the vertical slash sometimes play a crucial role for the applicability of subsequent inferences involving the directional slashes. A deductive system with such a radically hybrid property is unheard of and is surely unorthodox, and its formal underpinnings need to be investigated more closely, but we leave this task for future study.

CG.[4] As will become clear below, Oehrle's analysis illuminates the fundamental nature of scope-taking expressions by capturing the tight correlation between λ-binding in semantics and the corresponding binding of the 'gap' in the phonology, which is merely enforced by a bunch of coincidental stipulations in Montague's syncategorematic treatment of quantifying-in (which in turn is modelled in various ways in Lambek calculus-based variants of TLCG essentially carrying over the stipulative nature of Montague's solution as is). Unlike Montague's original treatment, in Oehrle's (1994) analysis the mismatch between linear position and scopal domain is a predicted *consequence* of hypothetical reasoning in the underlying logic.

The following analysis of the inverse-scope ($\forall > \exists$) reading of the sentence *Someone talked to everyone yesterday* illustrates how this works. We first hypothetically assume NPs in the surface positions in which the quantifiers appear. After the whole sentence is built up, withdrawing one of the hypotheses and binding the variables (in both phonology and semantics) via Vertical Slash Introduction produces a lambda-abstracted meaning/phonology pair that can be given to a quantifier as an argument. Combining the quantifier with such a lambda abstract has the effect that semantically the quantifier scopes over the whole expression but phonologically the string of the quantifier is inserted to the variable slot that is explicitly kept track of in the sentence's phonology via phonological lambda binding. The order in which the quantifiers are introduced in the derivation corresponds to their relative scope. Thus, in (20), since the universal quantifier in the object position is introduced in the derivation after the subject position quantifier, the inverse-scope interpretation is derived.

(20)

$$\dfrac{\lambda\sigma.\sigma(\mathsf{everyone});\ \forall_{\mathbf{person}};\ \mathrm{S}|(\mathrm{S}|\mathrm{NP}) \qquad \dfrac{\lambda\sigma.\sigma(\mathsf{someone});\ \exists_{\mathbf{person}};\ \mathrm{S}|(\mathrm{S}|\mathrm{NP}) \qquad \dfrac{\dfrac{\dfrac{\mathsf{talked}\circ\mathsf{to};\ \mathbf{talk\text{-}to};\ (\mathrm{NP}\backslash\mathrm{S})/\mathrm{NP} \qquad \left[\begin{smallmatrix}\varphi_1;\\ x;\mathrm{NP}\end{smallmatrix}\right]^1}{\mathsf{talked}\circ\mathsf{to}\circ\varphi_1;\ \mathbf{talk\text{-}to}(x);\mathrm{NP}\backslash\mathrm{S}}/\mathrm{E} \qquad \left[\begin{smallmatrix}\varphi_2;\\ y;\mathrm{NP}\end{smallmatrix}\right]^2}{\varphi_2\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1;\ \mathbf{talk\text{-}to}(x)(y);\mathrm{S}}\backslash\mathrm{E} \qquad \mathsf{yesterday};\ \mathbf{yest};\mathrm{S}\backslash\mathrm{S}}{\varphi_2\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1\circ\mathsf{yesterday};\ \mathbf{yest}(\mathbf{talk\text{-}to}(x)(y));\mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi_2.\varphi_2\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1\circ\mathsf{yesterday};\ \lambda y.\mathbf{yest}(\mathbf{talk\text{-}to}(x)(y));\mathrm{S}|\mathrm{NP}}|\mathrm{I}^2}{\mathsf{someone}\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1\circ\mathsf{yesterday};\ \exists_{\mathbf{person}}(\lambda y.\mathbf{yest}(\mathbf{talk\text{-}to}(x)(y)));\mathrm{S}}|\mathrm{E}}{\lambda\varphi_1.\mathsf{someone}\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1\circ\mathsf{yesterday};\ \lambda x.\exists_{\mathbf{person}}(\lambda y.\mathbf{yest}(\mathbf{talk\text{-}to}(x)(y)));\mathrm{S}|\mathrm{NP}}|\mathrm{I}^1}{\mathsf{someone}\circ\mathsf{talked}\circ\mathsf{to}\circ\mathsf{everyone}\circ\mathsf{yesterday};\ \forall_{\mathbf{person}}(\lambda x.\exists_{\mathbf{person}}(\lambda y.\mathbf{yest}(\mathbf{talk\text{-}to}(x)(y))));\mathrm{S}}|\mathrm{E}$$

As we have seen above, the use of the non-directional mode of implication enables a perspicuous treatment of scope, which essentially involves hypothetical reasoning

[4]Independent motivation for this technique comes from an analysis of extraction (see section 4.2 below) and Gapping (Kubota & Levine 2012).

for which the information of word order is in principle irrelevant. However, in a system with the non-directional mode of implication alone (such as Oehrle's (1994) original term-labelled calculus as well as Lambda Grammar (Muskens 2003), Abstract Categorial Grammar (de Groote 2001) and Linear Grammar (Pollard 2011)), coordination turns out to pose a significant challenge. This point was already noted by Muskens (2001) (see also Kubota (2010:section 3.2.1) for a more extended discussion on this point), but no satisfactory solution has been proposed for this problem in the literature until the present. The problem essentially comes from the fact that such frameworks recognize linguistic expressions with functional phonologies, and semantically functional expressions such as verbs come with functional phonologies which specify the way in which the phonologies of the arguments are combined with the phonology provided by the verb. However, the standard analysis of coordination in CG rests on the assumption that the phonologies of all linguistic expressions are simply strings (and not functions over strings). Thus, even the simplest cases of coordination involving lexical verbs such as *John saw and kissed Mary* turns out to be extremely problematic for such approaches. To analyze the coordination of expressions with functional phonologies (with different arity), we need to posit a series of prosodically distinct entries of *and* (and *or*). Working this out technically is by no means a trivial issue, but the more serious problem is that non-directional variants of CG do not distinguish between left-looking from right-looking functors. Thus, with an entry for *and* which coordinates type $str \rightarrow str$ conjuncts (which is needed for the analysis of Right-Node Raising examples like *John saw, and Bill hit, Peter*), an S|NP expression *John saw* __ will be able to be coordinated with an expression __ *hit Peter* of the same type, resulting in quite serious overgeneration.

Since an analysis of (non-standard) coordination is central to our account of the interaction between coordination and scope, rather than attempting to solve this recalcitrant theoretical problem within a non-directional variant of CG, we propose to extend such a system by recognizing both the directional and non-directional modes of implication within a single calculus, incorporating the insights of the Lambek-inspired tradition of TLCG. Since our calculus recognizes the Introduction and Elimination rules for forward and backward slashes from standard TLCG, the analysis of NCC in it straightforwardly carries over to the present setup, which we illustrate below.

With the Introduction and Elimination rules for directional slashes, the analysis of nonconstituent coordination originally due to Dowty's (1988) and Steedman's (1985) CCG analyses and later incorporated in TLCG by Morrill (1994) carries over to the present setup straightforwardly. The idea behind this analysis is essentially that, in the setup of TLCG, hypothetical reasoning with forward and backward slashes enables us to reanalyze any substring of a sentence as a full-fledged 'constituent' (with an appropriate, higher-order semantic interpretation) that has the right combinatorial property such that it returns a sentence when it combines with the rest of the sentence. The derivation in (22) shows how the string *Bill the book* in (21) is reanalyzed as such a non-standard constituent.

(21) Mary gave [Bill the book] and [John the record].

(22)

$$\dfrac{\dfrac{\dfrac{[\varphi;\ f;\ \mathrm{VP/NP/NP}\ ]^1 \quad \mathsf{bill};\ \mathbf{b};\ \mathrm{NP}}{\varphi \circ \mathsf{bill};\ f(\mathbf{b});\ \mathrm{VP/NP}}\,/\mathrm{E} \quad \mathsf{the} \circ \mathsf{book};\ \textbf{the-book};\ \mathrm{NP}}{\varphi \circ \mathsf{bill} \circ \mathsf{the} \circ \mathsf{book};\ f(\mathbf{b})(\textbf{the-book});\ \mathrm{VP}}\,/\mathrm{E}}{\mathsf{bill} \circ \mathsf{the} \circ \mathsf{book};\ \lambda f.f(\mathbf{b})(\textbf{the-book});\ (\mathrm{VP/NP/NP})\backslash\mathrm{VP}}\,\backslash\mathrm{I}^1$$

The key step in the above derivation is the hypothetical assumption of a ditransitive verb. This hypothetical verb combines with the two object NPs *Bill* and *the book* just like ordinary ditransitive verbs and forms a VP. Then, the hypothesis is withdrawn to assign the category $(\mathrm{VP/NP/NP})\backslash\mathrm{VP}$ to the string *Bill the book*. Intuitively, this is saying that this string is something that becomes a VP if it finds a ditransitive verb to its left. Once this complex category is assigned to the string *Bill the book*, the rest just involves coordinating this non-standard constituent with another constituent with the same syntactic category via the standard generalized conjunction category for the coordinator *and* (where $\sqcap$ denotes generalized conjunction *a la* Partee & Rooth (1983)), and then putting the whole coordinated expression together with the verb and the subject NP as in (23).

(23)

$$\dfrac{\mathsf{mary};\ \mathbf{m};\ \mathrm{NP} \qquad \dfrac{\mathsf{gave};\ \textbf{give};\ \mathrm{VP/NP/NP} \qquad \dfrac{\mathsf{bill} \circ \mathsf{the} \circ \mathsf{book};\ \lambda f.f(\mathbf{b})(\textbf{the-book});\ (\mathrm{VP/NP/NP})\backslash\mathrm{VP} \qquad \dfrac{\mathsf{and};\ \lambda V\lambda W.W \sqcap V;\ (X\backslash X)/X \qquad \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \lambda f.f(\mathbf{j})(\textbf{the-record});\ (\mathrm{VP/NP/NP})\backslash\mathrm{VP}}{\mathsf{and} \circ \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \lambda W.W \sqcap \lambda f.f(\mathbf{j})(\textbf{the-record});\ ((\mathrm{VP/NP/NP})\backslash\mathrm{VP})\backslash((\mathrm{VP/NP/NP})\backslash\mathrm{VP})}\,/\mathrm{E}}{\mathsf{bill} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \lambda f.f(\mathbf{b})(\textbf{the-book}) \sqcap \lambda f.f(\mathbf{j})(\textbf{the-record});\ (\mathrm{VP/NP/NP})\backslash\mathrm{VP}}\,\backslash\mathrm{E}}{\mathsf{gave} \circ \mathsf{bill} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \textbf{give}(\mathbf{b})(\textbf{the-book}) \sqcap \textbf{give}(\mathbf{j})(\textbf{the-record});\ \mathrm{VP}}\,\backslash\mathrm{E}}{\mathsf{mary} \circ \mathsf{gave} \circ \mathsf{bill} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \textbf{give}(\mathbf{b})(\textbf{the-book})(\mathbf{m}) \wedge \textbf{give}(\mathbf{j})(\textbf{the-record})(\mathbf{m});\ \mathrm{S}}\,\backslash\mathrm{E}$$

Thus, positing both directional and non-directional modes of implication within a single calculus enables straightforward analyses of two kinds of major empirical phenomena (i.e. coordination and scopal expressions) that pose problems for previous variants of CG, which recognizes only one of these two types of implication. But the real strength of the hybrid implication architecture of the present framework becomes fully apparent in the analyses of phenomena like those discussed in section 2 in which coordination interacts with scopal expressions. These phenomena call for a system in which the mechanisms dealing with word order-related inferences (for coordination) and those dealing with order-insensitive reasoning (for scope) interact with one another systematically. The present framework provides precisely such an architecture, and we will see in the next section that the proper analysis of these more complex cases in fact falls out straightforwardly from the hybrid architecture of the present framework.

4 Coordination in Hybrid Type-Logical Categorial Grammar

4.1 NCC and symmetrical predicates

For the analysis of symmetrical predicates, we adopt the proposal by Barker (2007) in terms of *parasitic scope*. Pollard (2009) (described in Pollard & Smith (to appear)) implements this analysis in a term-labelling system with phonological λ-abstraction like Oehrle's (1994) setup. We adopt this implementation in our account of the interaction between NCC and symmetrical predicates. Barker's analysis of symmetrical predicates involves the following three elements as the key components in the semantic analysis of symmetrical predicates such as *(the) same*:[5]

(i) a property provided by the 'head noun' modified by the symmetrical predicate

(ii) a sum-denoting expression

(iii) a relation provided by the rest of the sentence (i.e. a structure obtained by abstracting over the NP containing the symmetrical predicate and the sum-denoting expression from the whole sentence)

For an example like (24), (i)–(iii) above are instantiated by the noun *waiter*, the coordinated NP *John and Bill*, and the transitive verb *served*, respectively.

(24) The same waiter served John and Bill.

The semantic contribution of *(the) same* on the internal reading is to assert the existence of some unique waiter (i.e. an individual satisfying the property (i)) such that the x-served-y relation (i.e. the relation provided by (iii)) holds between that individual and each atomic subpart of the plurality of John and Bill (i.e. the sum denoted by (ii)). Technically, the denotation of the symmetrical predicate *the same* is formulated as in (25), as a relation between its three semantic arguments (i)–(iii):

[5] We adopt this three-place function analysis of *same* due to Barker (2007) here for expository purposes. Note, however, that this analysis does not easily generalize to cases involving multiple occurrences of *same* invoking the internal reading with respect to the same plural entity at the same time, as exemplified by the following example:

(i) John and Bill bought the same book at the same store (on the same day ...).

In Kubota & Levine (2013), we provide a more complete analysis of symmetrical predicates which can deal with iterated *same* examples like (i). This latter analysis is superior to the one we adopt here also in that it extends straightforwardly to related expressions such as 'respective' and summative predicates observed above in (5) and (6), and captures the complex (yet systematic) interactions between these three classes of phenomena in a uniform manner.

(25) $\lambda P \lambda Q \lambda X.\exists y[P(y) \wedge \forall x.x <_a X \rightarrow Q(y)(x)]$

Syntactically, the key steps of the derivation involves forming constituents that respectively provide the three semantic arguments to the symmetrical predicate and combining them in the right order. Specifically, the relation-denoting expression can be obtained by abstracting over the positions corresponding to the NP containing *same* and the sum-denoting expression via successive applications of the Vertical Slash introduction rule (cf. the steps down to the fifth line of the rightmost chunk in (26)). This creates a doubly abstracted proposition of syntactic category $(\mathrm{S}|\mathrm{NP})|\mathrm{NP}$ and this doubly abstracted proposition is then given as an argument to the expression *the same*, together with the other two arguments. Crucially, the phonology of the symmetrical predicate is specified as a higher-order functional phonological term which, if it takes the three arguments it semantically requires in the right order, produces the string of words corresponding to the surface form of the sentence.

(26)

$$\dfrac{\dfrac{\left[\begin{matrix}\psi; \\ y; \mathrm{NP}\end{matrix}\right]^2 \quad \dfrac{\begin{matrix}\mathsf{served}; \\ \mathbf{serve}; \mathrm{VP/NP}\end{matrix} \quad \left[\begin{matrix}\varphi; \\ x; \mathrm{NP}\end{matrix}\right]^1}{\begin{matrix}\mathsf{served} \circ \varphi; \\ \mathbf{serve}(x); \mathrm{VP}\end{matrix}}/\mathrm{E}}{\psi \circ \mathsf{served} \circ \varphi; \mathbf{serve}(x)(y); \mathrm{S}}\backslash\mathrm{E}$$

$$\dfrac{\psi \circ \mathsf{served} \circ \varphi; \mathbf{serve}(x)(y); \mathrm{S}}{\begin{matrix}\lambda\varphi.\psi \circ \mathsf{served} \circ \varphi; \\ \lambda x.\mathbf{serve}(x)(y); \mathrm{S}|\mathrm{NP}\end{matrix}}|\mathrm{I}^1 \qquad \dfrac{\begin{matrix}\lambda\varphi.\psi \circ \mathsf{served} \circ \varphi; \\ \lambda x.\mathbf{serve}(x)(y); \mathrm{S}|\mathrm{NP}\end{matrix}}{\begin{matrix}\lambda\psi\lambda\varphi.\psi \circ \mathsf{served} \circ \varphi; \\ \lambda x.\mathbf{serve}(x)(y); \mathrm{S}|\mathrm{NP}|\mathrm{NP}\end{matrix}}|\mathrm{I}^2$$

$$\dfrac{\begin{matrix}\lambda\psi\lambda\sigma.\sigma(\mathsf{the} \circ \mathsf{same} \circ \psi); \\ \mathbf{same}; \\ ((\mathrm{S}|X)|((\mathrm{S}|X)|\mathrm{NP}))|\mathrm{N}\end{matrix} \quad \begin{matrix}\mathsf{waiter}; \\ \mathbf{waiter}; \\ \mathrm{N}\end{matrix}}{\begin{matrix}\lambda\sigma.\sigma(\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter}); \\ \mathbf{same}(\mathbf{waiter}); (\mathrm{S}|X)|((\mathrm{S}|X)|\mathrm{NP})\end{matrix}}|\mathrm{E}$$

$$\dfrac{\begin{matrix}\lambda\sigma.\sigma(\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter}); \\ \mathbf{same}(\mathbf{waiter}); (\mathrm{S}|X)|((\mathrm{S}|X)|\mathrm{NP})\end{matrix} \quad \begin{matrix}\lambda\psi\lambda\varphi.\psi \circ \mathsf{served} \circ \varphi; \\ \lambda x.\mathbf{serve}(x)(y); \mathrm{S}|\mathrm{NP}|\mathrm{NP}\end{matrix}}{\begin{matrix}\lambda\varphi.\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter} \circ \mathsf{served} \circ \varphi; \\ \mathbf{same}(\mathbf{waiter})(\lambda y \lambda x.\mathbf{serve}(x)(y)); \mathrm{S}|\mathrm{NP}\end{matrix}}|\mathrm{E}$$

$$\dfrac{\begin{matrix}\vdots \quad \vdots \\ \mathsf{john} \circ \mathsf{and} \circ \mathsf{bill}; \\ \mathbf{j} \oplus \mathbf{b}; \mathrm{NP}\end{matrix} \quad \begin{matrix}\lambda\varphi.\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter} \circ \mathsf{served} \circ \varphi; \\ \mathbf{same}(\mathbf{waiter})(\lambda y \lambda x.\mathbf{serve}(x)(y)); \mathrm{S}|\mathrm{NP}\end{matrix}}{\begin{matrix}\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter} \circ \mathsf{served} \circ \mathsf{john} \circ \mathsf{and} \circ \mathsf{bill}; \\ \mathbf{same}(\mathbf{waiter})(\lambda y \lambda x.\mathbf{serve}(x)(y))(\mathbf{j} \oplus \mathbf{b}); \mathrm{S}\end{matrix}}|\mathrm{E}$$

(In the above derivation, the variables abstracted over happen to correspond to subject and object arguments of the transitive verb, and thus these two abstraction steps might appear to be (semantically) superfluous. However, as will become clear in more complicated derivations below, this is not generally the case, and we need the two abstraction steps for the vertical slash in order to provide a general treatment of symmetrical predicates.) By unpacking the final translation for the sentence derived in (26), we get the following truth conditions:

(27) $\mathbf{same}(\mathbf{waiter})(\lambda y \lambda x.\mathbf{serve}(x)(y))(\mathbf{j} \oplus \mathbf{b}) = \exists y[\mathbf{waiter}(y) \wedge \forall x.x <_a \mathbf{j} \oplus \mathbf{b} \rightarrow \mathbf{serve}(x)(y)]$

This says that there is a waiter such that for both of John and Bill, it is the case that that waiter served him. In other words, there is a unique waiter who served both John and Bill. This correctly captures the meaning of the sentence on the relevant reading.

This analysis of symmetrical predicates by Barker implemented in a system with phonological lambda abstraction interacts straightforwardly with the direct licensing analysis of NCC from directional CG to assign the right interpretations for sentences like (28).

(28) Terry said the same thing to Robin on Thursday and to Leslie on Friday.

The crucial assumption that enables a straightforward extension of the analysis of symmetrical predicates for the simpler case involving coordination of simple NPs (denoting sums of type e objects) above to more complex cases like (28) is that in the lexical entry for *the same* in (27), the type of the sum-denoting expression (and, correspondingly, of the relation that takes subparts of that sum as one of its arguments) is polymorphic. Specifically, in the case of (28), the sum involved is a sum of higher-order semantic objects of type $e \to (e \to e \to e \to t) \to e \to t$. Other than this slight complication in the semantic type, the function of the symmetrical predicate *the same* is the same as in the previous case: it asserts that an identical relation holds between each subpart of this sum and some unique entity satisfying the descriptive content provided by the nominal head that the symmetrical predicate combines with. Thus, the analysis is essentially parallel to the simpler case involving coordination of ordinary NPs in (24). The derivation for (28) is given in (29).

(29)

$$\dfrac{\dfrac{\dfrac{\begin{array}{c}\text{terry};\\ \mathbf{t}; \mathrm{NP}\end{array} \quad \dfrac{\begin{array}{c}\text{said};\\ \mathbf{say}; \mathrm{VP/PP/NP}\end{array} \quad \dfrac{[\varphi_1; x; \mathrm{NP}]^1 \quad [\varphi_2; f; \mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash \mathrm{VP}]^2}{\varphi_1 \circ \varphi_2; f(x); (\mathrm{VP/PP/NP})\backslash \mathrm{VP}}\backslash\mathrm{E}}{\text{said} \circ \varphi_1 \circ \varphi_2; f(x)(\mathbf{say}); \mathrm{VP}}\backslash\mathrm{E}}{\text{terry} \circ \text{said} \circ \varphi_1 \circ \varphi_2; f(x)(\mathbf{say})(\mathbf{t}); \mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi_2.\text{terry} \circ \text{said} \circ \varphi_1 \circ \varphi_2; \lambda x.f(x)(\mathbf{say})(\mathbf{t}); \mathrm{S}|(\mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash \mathrm{VP})}|\mathrm{I}^1}{\lambda\varphi_2\lambda\varphi_1.\text{terry} \circ \text{said} \circ \varphi_1 \circ \varphi_2; \lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t}); \mathrm{S}|(\mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash \mathrm{VP})|\mathrm{NP}}|\mathrm{I}^2$$

$$\dfrac{\begin{array}{c}\vdots \quad \vdots\\ \text{to} \circ \text{robin} \circ \text{on} \circ \text{thursday};\\ \lambda x\lambda P.\mathbf{onTh}(P(x)(\mathbf{r}));\\ \mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash \mathrm{VP}\end{array} \quad \dfrac{\begin{array}{c}\text{and};\\ \lambda X\lambda Y.X \oplus Y;\\ (X\backslash X)/X\end{array} \quad \begin{array}{c}\vdots \quad \vdots\\ \text{to} \circ \text{leslie} \circ \text{on} \circ \text{friday};\\ \lambda x\lambda P.\mathbf{onFr}(P(x)(\mathbf{l}));\\ \mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash \mathrm{VP}\end{array}}{\begin{array}{c}\text{and} \circ \text{to} \circ \text{robin} \circ \text{on} \circ \text{thursday};\\ \lambda X.X \oplus [\lambda x\lambda P.\mathbf{onFr}(P(x)(\mathbf{l}))];\\ (\mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash \mathrm{VP})\backslash(\mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash \mathrm{VP})\end{array}}/\mathrm{E}}{\begin{array}{c}\text{to} \circ \text{robin} \circ \text{on} \circ \text{thursday} \circ \text{and} \circ \text{to} \circ \text{leslie} \circ \text{on} \circ \text{friday};\\ \lambda x\lambda P\mathbf{onFr}(P(x)(\mathbf{l})) \oplus \lambda x\lambda P.\mathbf{onTh}(P(x)(\mathbf{r})); \mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash \mathrm{VP}\end{array}}\backslash\mathrm{E}$$

$$\dfrac{\dfrac{\begin{array}{c}\vdots\quad\vdots\\ \mathsf{to}\circ\mathsf{robin}\circ\mathsf{on}\circ\mathsf{thursday}\circ\\ \mathsf{and}\circ\mathsf{to}\circ\mathsf{leslie}\circ\mathsf{on}\circ\mathsf{friday};\\ [\lambda x\lambda P\lambda z.\mathbf{onFr}(P(x)(\mathbf{l}))(z)]\\ \oplus[\lambda x\lambda P\lambda z.\mathbf{onTh}(P(x)(\mathbf{r}))(z)];\\ \mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash\mathrm{VP}\end{array}\qquad\dfrac{\begin{array}{ll}\lambda\sigma.\sigma(\mathsf{the}\circ & \vdots\quad\vdots\\ \mathsf{same}\circ\mathsf{thing}); & \lambda\varphi_1\lambda\varphi_2.\mathsf{terry}\circ\mathsf{said}\circ\varphi_1\circ\varphi_2;\\ \mathbf{same}(\mathbf{thing}); & \lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t});\\ (\mathrm{S|X})|(\mathrm{S|X|NP}) & \mathrm{S|(NP\backslash(VP/PP/NP)\backslash VP)|NP}\end{array}}{\lambda\varphi_2.\mathsf{terry}\circ\mathsf{said}\circ\mathsf{the}\circ\mathsf{same}\circ\mathsf{thing}\circ\varphi_2;\ \mathbf{same}(\mathbf{thing})(\lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t}));\ \mathrm{S|(NP\backslash(VP/PP/NP)\backslash VP)}}\,|\mathrm{E}}{\begin{array}{l}\mathsf{terry}\circ\mathsf{said}\circ\mathsf{the}\circ\mathsf{same}\circ\mathsf{thing}\circ\\ \mathsf{to}\circ\mathsf{robin}\circ\mathsf{on}\circ\mathsf{thursday}\circ\mathsf{and}\circ\mathsf{to}\circ\mathsf{leslie}\circ\mathsf{on}\circ\mathsf{friday};\\ \mathbf{same}(\mathbf{thing})(\lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t}))([\lambda x\lambda P\lambda z.\mathbf{onFr}(P(x)(\mathbf{l}))(z)]\\ \oplus[\lambda x\lambda P\lambda z.\mathbf{onTh}(P(x)(\mathbf{r}))(z)]);\mathrm{S}\end{array}}\,|\mathrm{E}$$

What is crucial in this derivation is the part that derives the NCC involving the strings *to Robin on Thursday* and *to Leslie on Friday*. These strings are analyzed as (nonstandard) constituents via hypothetical reasoning in the same way as other examples above. They are then coordinated with the generalized sum meaning of *and* to form a (generalized) sum of type $e \to (e \to e \to e \to t) \to e \to t$ objects. The rest of the derivation involves creating a doubly-abstracted proposition by abstracting over the positions corresponding to the NP containing *same* and the (higher-order) sum-denoting expression and giving this proposition as an argument to the symmetrical predicate together with its other two arguments, namely, the (higher-order) sum derived above and the noun that provides the descriptive content for the unique entity involved. The translation for the whole sentence is unpacked and simplified in (30):

(30) $\mathbf{same}(\mathbf{thing})(\lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t}))([\lambda x\lambda P\lambda z.\mathbf{onFr}(P(x)(\mathbf{l}))(z)]\oplus[\lambda x\lambda P\lambda z.\mathbf{onTh}(P(x)(\mathbf{r}))(z)])$

$= \exists y[\mathbf{thing}(y) \wedge \forall R.R <_a [\lambda xLPLz.\mathbf{onFr}(P(x)(\mathbf{l}))(z)]$
$\oplus[\lambda xLP\lambda z.\mathbf{onTh}(P(x)(\mathbf{r}))(z)] \to R(y)(\mathbf{say})(\mathbf{t})]$

$= \exists y[\mathbf{thing}(y) \wedge \lambda x\lambda P\lambda z.[\mathbf{onFr}(P(x)(\mathbf{l}))(z)](y)(\mathbf{say})(\mathbf{t}) \wedge$
$\lambda x\lambda P\lambda z.[\mathbf{onTh}(P(x)(\mathbf{r}))(z)](y)(\mathbf{say})(\mathbf{t})]$

$= \exists y[\mathbf{thing}(y) \wedge \mathbf{onFr}(\mathbf{say}(y)(\mathbf{l}))(\mathbf{t}) \wedge \mathbf{onTh}(\mathbf{say}(y)(\mathbf{r}))(\mathbf{t})]$

This asserts the existence of some unique entity which was said by Terry both to Robin on Thursday and to Leslie on Friday. This correctly corresponds to the internal reading of the sentence where the matters communicated to the two people by Robin on different days are identical to each other.

4.2 UCC and extraction

The interaction between UCC and extraction also receives a straightforward solution in our approach. For the analysis of UCC, we adopt the proposal by Morrill (1994) and Bayer (1996) that involves extending the syntactic type system with the $\vee$ (join) connective. $\vee$ is a two place connective and the complex syntactic category $A \vee B$ intuitively means

that the linguistic expression that is assigned this category belongs to either category A or B. For the join connective, we posit the following two Introduction rules in our system:

(31) a. *Right Join Introduction*

$$\frac{\boldsymbol{a};\boldsymbol{f};A}{\boldsymbol{a};\boldsymbol{f};A\vee B}\vee\mathrm{I}$$

b. *Left Join Introduction*

$$\frac{\boldsymbol{a};\boldsymbol{f};B}{\boldsymbol{a};\boldsymbol{f};A\vee B}\vee\mathrm{I}$$

Intuitively, these rules say that if something is an A (or a B), then we are entitled to conclude a weaker statement that it is $A \vee B$ (i.e. A or B).

The key assumption in the Morrill/Bayer analysis of UCC is the specification of the copula given in (32):

(32) $\mathsf{is};\ \lambda f.f;\ \mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP})$

This says that *is* is looking for either an NP or an AP as its complement to become a VP. With the $\vee$-Introduction rule in (31), the derivation for a sentence in which the copula combines with an NP complement (without UCC) goes as follows:

(33)

$$\dfrac{\mathsf{pat};\mathrm{NP}\qquad\dfrac{\mathsf{is};\mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP})\qquad\dfrac{\mathsf{a}\circ\mathsf{republican};\mathrm{NP}}{\mathsf{a}\circ\mathsf{republican};\mathrm{NP}\vee\mathrm{AP}}\vee\mathrm{I}}{\mathsf{is}\circ\mathsf{a}\circ\mathsf{republican};\mathrm{VP}}/\mathrm{E}}{\mathsf{pat}\circ\mathsf{is}\circ\mathsf{a}\circ\mathsf{republican};\mathrm{S}}\backslash\mathrm{E}$$

The key point here is that, with $\vee$-Introduction, we can assign the category $\mathrm{NP}\vee\mathrm{AP}$ to the string *a Republican* and this satisfies the subcategorization requirement of the copula.

From this, it should already be clear how examples of UCC like *Pat is a Republican and proud of it* are derived. The derivation is given in (34).

(34)

$$\dfrac{\mathsf{pat};\mathrm{NP}\qquad\dfrac{\mathsf{is};\mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP})\qquad\dfrac{\dfrac{\mathsf{a}\circ\mathsf{republican};\mathrm{NP}}{\mathsf{a}\circ\mathsf{republican};\mathrm{NP}\vee\mathrm{AP}}\vee\mathrm{I}\qquad\dfrac{\mathsf{and};(X\backslash X)/X\qquad\dfrac{\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{AP}}{\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{NP}\vee\mathrm{AP}}\vee\mathrm{I}}{\mathsf{and}\circ\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};(\mathrm{NP}\vee\mathrm{AP})\backslash(\mathrm{NP}\vee\mathrm{AP})}/\mathrm{E}}{\mathsf{a}\circ\mathsf{republican}\circ\mathsf{and}\circ\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{NP}\vee\mathrm{AP}}\backslash\mathrm{E}}{\mathsf{is}\circ\mathsf{a}\circ\mathsf{republican}\circ\mathsf{and}\circ\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{VP}}/\mathrm{E}}{\mathsf{pat}\circ\mathsf{is}\circ\mathsf{a}\circ\mathsf{republican}\circ\mathsf{and}\circ\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{S}}\backslash\mathrm{E}$$

Here, both of the two conjuncts (i.e. the NP *a Republican* and the AP *proud of it*) are derived as NP$\vee$AP via $\vee$-Introduction. Then, with the standard generalized conjunction category for *and*, they are coordinated to form a larger constituent of category NP$\vee$AP. Since this category exactly matches the category that the copula is looking for as its argument, this UCC constituent can be directly combined with the copula via Slash Elimination to complete the derivation. Unlike the ellipsis-based approach of the kind advocated in the LBE

literature, the Morrill/Bayer analysis of UCC in CG treats strings like *a Republican and proud of it* as full-fledged surface constituents without any deletion operation or dummy syntactic head of any kind. As will become clear below, this turns out to be crucial in the analysis of the interactions between UCC and extraction.

For the analysis of extraction, we adopt the proposal by Muskens (2003) that exploits the order-insensitive nature of the non-directional mode of implication (i.e. our vertical slash). In the TLCG literature, the treatment of extraction has long been known as a problematic issue. Essentially, the problem is that modelling extraction by means of forward and backward slashes makes it difficult treat cases of extraction from non-peripheral positions. This is because Slash Introduction can apply for the forward and backward slashes only when (the phonology of) the hypothesis appears at a peripheral position. (An analogous problem is found with CCG, which deals with non-peripheral extraction via order-disrupting, non-harmonic function composition rules.) Various mechanisms have been proposed in the TLCG literature to overcome this problem, but they all involve significant complications in the mapping between syntax and surface morpho-phonology. Muskens's proposal is unique in that it solves this problem by directly representing the phonology of gapped sentences via a higher-order functional phonological term, using a mechanism independently needed in the grammar, namely, λ-binding in phonology. This simplifies the treatment of filler-gap dependency in the CG-based setup considerably.

The core idea of Muskens's (2003) approach to extraction involves analyzing (incomplete) sentences with gaps like *Kim likes* in the topicalization sentence in (35) as a sentence missing some expression somewhere inside, with hypothetical reasoning for the vertical slash, as in the derivation in (36).

(35) Bagels$_i$, Kim likes t_i

(36)

$$\dfrac{\dfrac{\mathsf{bagels};\ \mathbf{b};\ \mathrm{NP} \qquad \dfrac{\lambda\sigma\lambda\varphi.\varphi\circ\sigma(\boldsymbol{\epsilon});\ \lambda f.f;\ (\mathrm{S}|\mathrm{X})|(\mathrm{S}|\mathrm{X}) \qquad \dfrac{\dfrac{\mathsf{kim};\ \mathbf{k};\ \mathrm{NP} \qquad \dfrac{\mathsf{likes};\ \mathbf{like};\ (\mathrm{NP}\backslash\mathrm{S})/\mathrm{NP} \qquad \left[\begin{smallmatrix}\varphi;\\ x;\ \mathrm{NP}\end{smallmatrix}\right]^1}{\mathsf{likes}\circ\varphi;\ \mathbf{like}(x);\ \mathrm{NP}\backslash\mathrm{S}}/\mathrm{E}}{\mathsf{kim}\circ\mathsf{likes}\circ\varphi;\ \mathbf{like}(x)(\mathbf{k});\ \mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi.\mathsf{kim}\circ\mathsf{likes}\circ\varphi;\ \lambda x.\mathbf{like}(x)(\mathbf{k});\ \mathrm{S}|\mathrm{NP}}|\mathrm{I}^1}{\lambda\varphi.\varphi\circ\mathsf{kim}\circ\mathsf{likes};\ \lambda x.\mathbf{like}(x)(\mathbf{k});\ \mathrm{S}|\mathrm{NP}}}{\mathsf{bagels}\circ\mathsf{kim}\circ\mathsf{likes};\ \lambda x.\mathbf{like}(\mathbf{b})(\mathbf{k});\ \mathrm{S}}|\mathrm{E}$$

In (36), an NP is hypothesized in the object position of the transitive verb, and by withdrawing this hypothesis after the whole sentence is built, the position of the gap in the whole string is explicitly represented by the phonology of the hypothesized NP bound by the λ-operator, namely, the variable φ. (Note also that this gapped sentence is assigned the right meaning, that is, the property of being an object that Kim likes, with lambda abstraction of the variable x over the meaning of the whole sentence.) Since hypothetical reasoning for the vertical slash can be carried out regardless of the position of the variable in the surface string (unlike for the directionality-sensitive, forward and backward slashes), this approach can treat filler-gap dependency in a fully general manner wherever the gap appears within

the sentence.

For topicalization, the filler that corresponds to the gap appears in a position immediately to the left of the gapped sentence in the surface string. There is thus a mismatch between the surface form of the sentence and the phonology of the gapped constituent (which takes some string as an argument and embeds it in the original gap site), and this mismatch is mediated by the following phonologically empty topicalization operator:

(37) $\lambda\sigma\lambda\varphi.\varphi \circ \sigma(\boldsymbol{\epsilon});\ \lambda f.f;\ (\mathrm{S}|\mathrm{X})|(\mathrm{S}|\mathrm{X})$

The topicalization operator in (37) does not have any effect on either syntactic category or semantics; it only changes the phonology of the expression it combines with in such a way that, by combining with this topicalization operator, an empty string $\boldsymbol{\epsilon}$ is embedded in the original gap site and the host sentence now concatenates the phonology of its argument (i.e. the filler) immediately to the left of its own phonology.

With this analysis of topicalization, sentences like (38) in which a coordinate structure involving UCC gets topicalized can be analyzed as in (39).

(38) [(Both) poor and a Republican]$_i$ you can't possibly be t_i.

(39)

$$\dfrac{\begin{array}{lll}\vdots\ \vdots & \lambda\sigma\lambda\varphi.\varphi\circ\sigma(\boldsymbol{\epsilon});\ (\mathrm{S}|\mathrm{X})|(\mathrm{S}|\mathrm{X}) & \dfrac{\dfrac{\mathsf{you};\ \mathrm{NP} \qquad \dfrac{\mathsf{can't}\circ\mathsf{be};\ \mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP}) \qquad [\varphi;\mathrm{NP}\vee\mathrm{AP}]^1}{\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{VP}}/\mathrm{E}}{\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi.\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{S}|(\mathrm{NP}\vee\mathrm{AP})}|\mathrm{I}^1\\ \mathsf{both}\circ\mathsf{poor}\circ\mathsf{and}\circ\mathsf{a}\circ\mathsf{republican};\ \mathrm{NP}\vee\mathrm{AP} & & \lambda\varphi.\varphi\circ\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be};\mathrm{S}|(\mathrm{NP}\vee\mathrm{AP})\end{array}}{\mathsf{both}\circ\mathsf{poor}\circ\mathsf{and}\circ\mathsf{a}\circ\mathsf{republican}\circ\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be};\mathrm{S}}|\mathrm{E}\ |\mathrm{E}$$

The key step in this derivation is the hypothetical assumption of an expression of category $\mathrm{NP}\vee\mathrm{AP}$ in the gap position. Via hypothetical reasoning for the vertical slash, a gapped sentence of category $\mathrm{S}|(\mathrm{NP}\vee\mathrm{AP})$ can then be derived, which is missing an expression of category $\mathrm{NP}\vee\mathrm{AP}$, i.e., the complement of the copula. As already shown in the derivation of a simpler UCC sentence above in (34), the string *both poor and a Republican*, which appears in the filler position in (39), can be assigned the category $\mathrm{NP}\vee\mathrm{AP}$. Thus, the gap and the filler match in syntactic category and the two can be combined by means of the topicalization operator in the same way as the previous NP topicalization example in (36).

The (truth-conditional) meaning of a topicalization sentence is obtained simply by substituting the meaning of the filler in the gap position of the host sentence. Thus, in (39), the correct meaning is assigned to the whole sentence in which a conjunction of the two properties is predicated of the subject under the scope of modal and negation.

The analysis of pseudo-cleft is similarly straightforward. Again, the key assumption is the treatment of the gapped sentence with the vertical slash. The gapped sentence is derived in the category $\mathrm{S}|X$, a sentence missing an X somewhere inside. We assign the syntactic category $X|(\mathrm{S}|X)$ to the word *what*, which combines with this gapped sentence and forms the constituent that occupies the precopular position. As illustrated in the following derivation, by assigning this category to *what*, the constituent in the precopular position ends up having the same syntactic category as the gap. Then, with the syntactic category $(X\backslash\mathrm{S})/X$ for the copula, which identifies the categories of its left-hand and right-hand arguments, it follows that the syntactic categories of the gap and the postcopular expression are required to match with one another, capturing the basic syntactic properties of the pseudo-cleft construction. The derivation for the sentence *What Robin wanted was a textbook* is shown in (40):

(40)

$$\dfrac{\dfrac{\lambda\sigma.\mathsf{what}\circ\sigma(\boldsymbol{\epsilon});\ \mathrm{X}|(\mathrm{S}|\mathrm{X}) \qquad \dfrac{\dfrac{\mathsf{robin};\ \mathrm{NP} \qquad \dfrac{\mathsf{wanted};\ (\mathrm{NP}\backslash\mathrm{S})/\mathrm{NP} \qquad [\varphi;\mathrm{NP}]^1}{\mathsf{wanted}\circ\varphi;\mathrm{NP}\backslash\mathrm{S}}/\mathrm{E}}{\mathsf{robin}\circ\mathsf{wanted}\circ\varphi;\mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi.\mathsf{robin}\circ\mathsf{wanted}\circ\varphi;\mathrm{S}|\mathrm{NP}}|\mathrm{I}^1}{\mathsf{what}\circ\mathsf{robin}\circ\mathsf{wanted};\mathrm{NP}}|\mathrm{E} \qquad \dfrac{\mathsf{was};\ (\mathrm{X}\backslash\mathrm{S})/\mathrm{X} \qquad \mathsf{a}\circ\mathsf{textbook};\ \mathrm{NP}}{\mathsf{was}\circ\mathsf{a}\circ\mathsf{textbook};\mathrm{NP}\backslash\mathrm{S}}/\mathrm{E}}{\mathsf{what}\circ\mathsf{robin}\circ\mathsf{wanted}\circ\mathsf{was}\circ\mathsf{a}\circ\mathsf{textbook};\mathrm{S}}|\mathrm{E}$$

As in the analysis of topicalization, the interaction between pseudocleft and UCC is straightforward. By assuming an expression of the complex syntactic category $\mathrm{NP}\vee\mathrm{AP}$ in the gap position, the precopular constituent headed by *what* is derived in the same category as the gap, namely, $\mathrm{NP}\vee\mathrm{AP}$. And then, with the polymorphic syntactic category for the copula, the syntactic category of the precopular and postcopular expressions (which is a UCC of category $\mathrm{NP}\vee\mathrm{AP}$) are identified with each other to complete the derivation. Again, with the assumption that the UCC category denotes a conjunction of two properties, the right semantics is assigned for (11a), where the conjunction of two properties scopes below the negation and the modal.

(41)

$$\dfrac{\dfrac{\lambda\sigma.\mathsf{what}\circ\sigma(\boldsymbol{\epsilon});\ \mathrm{X}|(\mathrm{S}|\mathrm{X}) \qquad \dfrac{\dfrac{\mathsf{you};\ \mathrm{NP} \qquad \dfrac{\mathsf{can't}\circ\mathsf{be};\ \mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP}) \qquad [\varphi;\mathrm{NP}\vee\mathrm{AP}]^1}{\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{VP}}/\mathrm{E}}{\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi.\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{S}|(\mathrm{NP}\vee\mathrm{AP})}|\mathrm{I}^1}{\mathsf{what}\circ\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be};\mathrm{NP}\vee\mathrm{AP}}|\mathrm{E} \qquad \dfrac{\mathsf{is};\ (\mathrm{X}\backslash\mathrm{S})/\mathrm{X} \qquad \begin{array}{c}\vdots\ \ \vdots\\ \mathsf{intelligentoand}\\ \mathsf{aofundamentalist};\\ \mathrm{NP}\vee\mathrm{AP}\end{array}}{\mathsf{is}\circ\mathsf{intelligent}\circ\mathsf{and}\circ\mathsf{a}\circ\mathsf{fundamentalist};\ (\mathrm{NP}\vee\mathrm{AP})\backslash\mathrm{S}}/\mathrm{E}}{\mathsf{what}\circ\mathsf{youcan't}\circ\mathsf{be}\circ\mathsf{is}\circ\mathsf{intelligent}\circ\mathsf{and}\circ\mathsf{a}\circ\mathsf{fundamentalist};\mathrm{S}}|\mathrm{E}$$

The join connective introduced above in the analysis of UCC has as its dual the meet connective. We will show in Appendix A that the use of this meet connective enables an analysis of examples of UCC with different subcategorization frames such as the following, first noted by Crysmann (2003) and taken to exemplify the superiority of an ellipsis-based analysis of coordination over the direct coordination analysis in CG.

(42) John gave Mary a book and to Peter a record.

4.3 Nominal head coordination

Finally, we analyze the apparent agreement mismatch between the determiner and the verb in nominal head coordination in examples like the following:

(43) That man and woman are arguing again.

Note first that the acceptability of this nominal head coordination pattern partly depends on the semantic/pragmatic properties of the conjoined nominals; as shown in the following examples, combinations of nouns that can naturally be thought of as forming pairs (e.g. *man and woman*, *boy and girl*, *table and chair*) can felicitously appear in this construction, whereas random combinations of nouns (e.g. *man and chair*, *table and boy*) that cannot naturally be construed as forming pairs are generally infelicitous in this construction.

(44) a. This $\left\{\begin{array}{c}\text{man and woman}\\ \text{boy and girl}\\ \text{table and chair}\end{array}\right\}$ are in perfect match.

b.??This $\left\{\begin{array}{c}\text{man and chair}\\ \text{table and boy}\end{array}\right\}$ are in perfect match.

We take this to indicate that the coordinated nominals such as *man and woman* in (43) are a special kind of pair-denoting nominals rather than simply an elliptical version of coordination of full-fledged NPs. (That is, if these examples were derived via ellipsis from coordination of full-fledged NPs, the acceptability contrast in (44) would be puzzling.)

The simplest way to capture this special property of nominal head coordination is to assume that the relevant semantic/pragmatic restriction is encoded in the definition of the covert operator that is responsible for converting the original meanings of such coordinated nominals to the appropriate pair-denoting meanings. With the generalized sum meaning for *and*, 'property sum' meanings of the following form are freely available for coordinated nominals like *man and woman*:

(45) $[\![\text{man and woman}]\!] = \lambda f \lambda g.[f \oplus g](\mathbf{man})(\mathbf{woman}) = \mathbf{man} \oplus \mathbf{woman}$

We posit a following phonologically empty pair-forming operator in (46) that takes such property sums as arguments and returns a property that holds of a pair of individuals just in case the pair in question each satisfy one of the two properties that are parts of the original property sum. The pair-forming operator additionally imposes a semantic/pragmatic restriction such that the original property sum constitutes a 'natural pair' (via the primitive predicate **natural-pair**, which we do not attempt to analyze further here).

(46) $\lambda\varphi.\varphi$; $\lambda P\lambda X.\mathbf{resp}(X)(P) \wedge \mathbf{natural\text{-}pair}(P)$; $X|X$

The pair-forming operator has as its core meaning the definition of the **resp** operator that is essentially identical to the one employed by Gawron & Kehler (2004) for the analysis of 'respective' sentences. The **resp** operator is defined as in (47):

(47) $\mathbf{resp}(X)(P) = 1$ iff
$\exists x, y[\mathbf{atom}(x) \wedge \mathbf{atom}(y) \wedge x \neq y \wedge X = x \oplus y \wedge \exists p, q <_a P[p \neq q \wedge p(x) \wedge q(y)]]$

$\mathbf{resp}(X)(P)$ is true of a sum of individuals X and a sum of properties P just in case there is a bijective relation between the set of individuals that are atomic subparts of X and the set of properties that are atomic subparts of P, such that for each such pair, the individual in question satisfies the property in question.

By applying the pair-forming operator to the property sum meaning of the nominal head coordination *man and woman*, we get the following set of pairs of individuals as output, which is a set of man-woman pairs:

(48) $[\![(46)]\!]([\![\text{man and woman}]\!])$
= $\lambda P\lambda X.[\mathbf{resp}(X)(P) \wedge \mathbf{natural\text{-}pair}(P)](\mathbf{man} \oplus \mathbf{woman})$
= $\lambda X.[\mathbf{resp}(X)(\mathbf{man} \oplus \mathbf{woman}) \wedge \mathbf{natural\text{-}pair}(\mathbf{man} \oplus \mathbf{woman})]$

On this approach, symmetrical modifiers such as *mutually incompatible*, which pose problems for the ellipsis-based analysis, can be treated simply as intersective modifiers that restrict the set of pairs (or groups) of individuals denoted by the head noun. Specifically, the meaning of *mutually incompatible* is given in (49), which takes a set of pairs (or groups) of individuals and imposes on these pairs (or groups) the further condition that they consist of members that are incompatible with each other.

(49) mutually ∘ incompatible; $\lambda P\lambda X.P(X) \wedge \mathbf{incompatible}(X)$; $\mathrm{N/N}$

With these assumptions about the denotations of coordinated nominal heads and symmetrical modifiers, the analysis for (14) goes as in (50):

(50)

$$\dfrac{\text{mutually} \circ \text{incompatible};\ \lambda P\lambda X.P(X) \wedge \mathbf{incmptbl}(X);\mathrm{N/N} \qquad \dfrac{\begin{array}{c}\vdots\ \ \vdots \\ \text{man} \circ \text{and} \circ \text{woman};\ \mathbf{man} \oplus \mathbf{woman};\mathrm{N} \qquad \lambda\varphi_1.\varphi_1;\ \lambda P\lambda X.\mathbf{resp}(X)(P);X|X\end{array}}{\text{man} \circ \text{and} \circ \text{woman};\ \lambda X.\mathbf{resp}(X)(\mathbf{man} \oplus \mathbf{woman});\mathrm{N}}\,|\mathrm{E}}{\text{mutually} \circ \text{incompatible} \circ \text{man} \circ \text{and} \circ \text{woman};\ \lambda X.\mathbf{resp}(X)(\mathbf{man} \oplus \mathbf{woman}) \wedge \mathbf{incmptbl}(X);\mathrm{N}}\,/\mathrm{E}$$

This denotes a set of man-woman pairs such that for each pair, the two individuals constituting the pair are incompatible with one another. The determiner *this* picks up a unique member from this set that is proximal to the speaker.

We take the apparent agreement mismatch between the determiner and the verb to receive a semantic account along the following lines. The singular agreement between the determiner and the coordinated nominal reflects the selectional restriction that the singular determiner imposes on the head noun such that the number of object(s) that satisfy the property denoted by the (coordinated) head noun is one. The plural agreement between the whole NP and the verb, on the other hand, reflects the number of object(s) (in terms of atomic individuals of type e) for which the verbal predicate holds. The man-woman pair that the subject NP denotes is semantically a sum of individuals (just like other plural NPs such as *John and Bill*), and thus triggers plural verb agreement. In short, there is an (apparent) agreement mismatch here since, for pair-denoting nominals, the determiner counts the number of pairs whereas the verb counts the number of members that constitute the pair(s).[6]

To summarize, here again, the right analysis that enables a systematic treatment of a wider range of facts involving symmetrical modifiers is not in terms of ellipsis, but one which directly assigns meanings to such apparently anomalous coordinate structures by means of (slight extensions of) independently motivated mechanisms of grammar such as the generalized sum meaning for *and* and the **resp** operator used in the analysis of 'respectively' sentences. Furthermore, the apparently anomalous agreement pattern, which at first sight appears to motivate an ellipsis-based analysis, receives a fully coherent account by means of an interaction between relevant syntactic and semantic factors.

5 Conclusion

In this paper, we discussed three cases in which non-standard coordination interacts with scopal expressions. The empirical generalization that emerges in these three cases is uniform: the scopal operator that appears outside the coordinate structure in the overt string always takes scope over the whole coordinate structure—in other words, the surface form of the sentence transparently reflects the relevant scopal relation. The null hypothesis in such a situation is that the syntactic constituency relevant for semantic interpretation mirrors this surface constituency between the coordinate structure and the scopal expression.

An ellipsis-based analysis of coordination like the LBE approach in the recent

[6]Heycock & Zamparelli (2005) argue against an analysis of data like (43) which posits an empty pair-forming operator (superficially) similar to our (46), by giving five reasons for rejecting such an analysis. All of their arguments crucially rest on the assumption that the inaudible pair-forming operator has exactly the same syntactic, semantic and morphological properties as the overt word *pair*. But note that such an assumption is dubious given that the distribution of expressions like *this man and woman (are)* is rather restricted (i.e. limited to cases that can be informally described by the notion of 'natural pair', as discussed in the main text) as compared to the overt noun *pair*, which does not come with any such restriction. For this reason, we take it that the facts discussed by Heycock and Zamparelli do not undermine our analysis.

HPSG literature goes wrong for this very reason. Specifically, in this type of approach, if the surface form of the sentence demands an analysis in which the scopal operator is part of the material that undergoes surface ellipsis, the default prediction is that the scopal operator takes scope *inside* each conjunct, but such readings are systematically lacking in all of the three cases considered above. For a similar discrepancy between the underlying combinatoric structure and the actual interpretation observed with generalized quantifiers, Beavers & Sag (2004) propose a mechanism called Optional Quantifier Merger, which specifically does away with the duplication of quantifier meanings from the final interpretation of the sentence on the condition that surface ellipsis takes place—effectively stipulating by fiat the effect that one would automatically get if the surface form of the sentence was directly assigned semantic interpretation without the ellipsis mechanism. Beavers & Sag's (2004) approach covers only the case of generalized quantifiers and it is not clear if it is extendable to other cases like those discussed in the present paper—in particular those involving symmetrical predicates, whose semantics is known to be more complex than that of ordinary generalized quantifiers (Keenan 1992; Barker 2007).

As we have discussed, CG offers a potentially very promising framework for analyzing these complex interactions between coordination and scopal expressions, given its transparent syntax-semantics interface and given its renowned 'direct coordination' analysis of non-standard coordination. However, the standard directional variants of CG is less than optimal for the treatment of scopal expressions due to the fact that the basic mode of implication dealing with syntactic combinatorics is inherently sensitive to word order. Thus, in order to analyze the interactions between scopal expressions and coordination in a fully general manner, we have chosen to extend a directional fragment, which is essentially a labelled deduction (re)formulation of the Lambek calculus, with a mechanism that deals with directionality-insensitive reasoning, incorporating the insight of Oehrle's (1994) term-labelled calculus for quantification. The resultant system recognizes both directional and non-directional modes of implication within a single calculus, and the two types of inference feed into one another freely. As we have shown above, this hybrid architecture of the present framework plays a crucial role in capturing the the empirical interactions between coordination (whose analysis involves inferences with the directional mode of implication) and scope-taking expressions (whose analysis involves the non-directional mode of implication). We thus conclude that the direct coordination analysis of non-standard coordination in CG is truly superior to an alternative that extensively relies on surface ellipsis like the LBE approach in the recent HPSG literature, but that the real empirical payoff of the direct coordination analysis becomes fully apparent only when it is embedded in a framework—like the one we have proposed in this paper—which can deal with complex yet systematic interactions between directional and non-directional modes of inference, each modelling the behaviors of different types of linguistic phenomena in a fully general manner.

References

ABBOTT, BARBARA. 1976. Right node raising as a test for constituenthood. *Linguistic Inquiry* 7.639–642.

BARKER, CHRIS. 2007. Parasitic scope. *Linguistics and Philosophy* 30.407–444.

BAYER, SAMUEL. 1996. The coordination of unlike categories. *Language* 72.579–616.

BEAVERS, JOHN, & IVAN A. SAG. 2004. Coordinate ellipsis and apparent non-constituent coordination. In *The Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Stefan Müller, 48–69, Stanford. CSLI.

BERNARDI, RAFFAELLA. 2002. *Reasoning with Polarity in Categorial Type Logic*. University of Utrecht dissertation. [Available at `http://www.inf.unibz.it/~bernardi/finalthesis.html`].

CARLSON, GREG N. 1987. Same and different: Some consequences for syntax and semantics. *Linguistics and Philosophy* 10.531–565.

CHAVES, RUI PEDRO. 2006. Coordination of unlikes without unlike categories. In *The Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Stefan Müller, 102–122, Stanford. CSLI Publications.

——. 2007. *Coordinate Structures - Constraint-based Syntax-Semantics Processing*. Portugal: University of Lisbon dissertation.

——. 2012. Conjunction, cumulation and respectively readings. *Journal of Linguistics* 48.297–344.

CRYSMANN, BERTHOLD. 2003. An asymmetric theory of peripheral sharing in HPSG: Conjunction reduction and coordination of unlikes. In *Proceedings of Formal Grammar 2003*, ed. by Gerhard Jäger, Paola Monachesi, Gerald Penn, & Shuly Wintner, 47–62. Available at http://cs.haifa.ac.il/~shuly/fg03/.

DE GROOTE, PHILIPPE. 2001. Towards abstract categorial grammars. In *Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference*, 148–155.

DOWTY, DAVID. 1988. Type raising, functional composition, and non-constituent conjunction. In *Categorial Grammars and Natural Language Structures*, ed. by Richard T. Oehrle, Emmon Bach, & Deirdre Wheeler, 153–198. Dordrecht: D. Reidel Publishing Company.

GAWRON, JEAN MARK, & ANDREW KEHLER. 2004. The semantics of respective readings, conjunction, and filler-gap dependencies. *Linguistics and Philosophy* 27.169–207.

GAZDAR, GERALD. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12.155–184.

HEYCOCK, CAROLINE, & ROBERTO ZAMPARELLI. 2005. Friends and colleagues: Plurality, coordination, and the structure of dp. *Natural Language Semantics* 13.201–270.

JACKENDOFF, RAY. 1977. *X-bar Syntax: A Study of Phrase Structure*. Cambridge, MA, USA: MIT Press.

KATHOL, ANDREAS. 1995. *Linearization-Based German Syntax*. Columbus: Ohio State University dissertation.

KEENAN, EDWARD L. 1992. Beyond the Frege boundary. *Linguistics and Philosophy* 15.199–221.

KUBOTA, YUSUKE. 2010. *(In)flexibility of Constituency in Japanese in Multi-Modal Categorial Grammar with Structured Phonology*. The Ohio State University dissertation.

——. to appear. The logic of complex predicates: A deductive synthesis of 'argument sharing' and 'verb raising'. To appear in *Natural Language and Linguistic Theory*.

——, & ROBERT LEVINE. 2012. Gapping as like-category coordination. In *Logical Aspects of Computational Linguistics: 7th International Conference*, ed. by Denis Béchet & Alexander Dikovsky, 135–150. Springer.

——, & ROBERT LEVINE. 2013. Against ellipsis: Arguments for the direct licensing of 'non-canonical' coordinations. MS., University of Tokyo and Ohio State University.

——, & CARL POLLARD. 2010. Phonological interpretation into preordered algebras. In *The Mathematics of Language: 10th and 11th Biennial Conference*, ed. by Christian Ebert, Gerhard Jäger, & Jens Michaelis, 200–209. Springer.

LAMBEK, JOACHIM. 1958. The mathematics of sentence structure. *American Mathematical Monthly* 65.154–170.

LEVINE, ROBERT. 2011. Linerarization and its discontents. In *The Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Stefan Müller, 126–146, Stanford. CSLI Publications.

MOORTGAT, MICHAEL. 1997. Categorial Type Logics. In *Handbook of Logic and Language*, ed. by Johan van Benthem & Alice ter Meulen, 93–177. Amsterdam: Elsevier.

——, & RICHARD T. OEHRLE. 1994. Adjacency, dependence, and order. In *Proceedings of the Ninth Amsterdam Colloquium*, ed. by Paul Dekker & Martin Stokhof, 447–466, Universiteit van Amsterdam. Instituut voor Taal, Logica, en Informatica.

MORRILL, GLYN, & TERESA SOLIAS. 1993. Tuples, discontinuity, and gapping in categorial grammar. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics*, 287–297, Morristown, NJ. Association for Computational Linguistics.

MORRILL, GLYN V. 1994. *Type Logical Grammar: Categorial Logic of Signs*. Dordrecht: Kluwer Academic Publishers.

MUSKENS, REINHARD. 2001. Categorial Grammar and Lexical-Functional Grammar. In *The Proceedings of the LFG '01 Conference*, ed. by Miriam Butt & Tracy Holloway King, University of Hong Kong.

——. 2003. Language, lambdas, and logic. In *Resource Sensitivity in Binding and Anaphora*, ed. by Geert-Jan Kruijff & Richard Oehrle, Studies in Linguistics and Philosophy, 23–54. Kluwer.

——. 2007. Separating syntax and combinatorics in categorial grammar. *Research on Language and Computation* 5.267–285.

OEHRLE, RICHARD T. 1994. Term-labeled categorial type systems. *Linguistics and Philosophy* 17.633–678.

PARTEE, BARBARA, & MATS ROOTH. 1983. Generalized quantifiers and type ambiguity. In *Meaning, Use, and Interpretation of Language*, ed. by Rainer Bäuerle, Christoph Schwarze, & Arnim von Stechow, 361–383. Berlin: Walter de Gruyter.

POLLARD, CARL. 2009. Parasitic scope in categorial grammar with φ-labelling. Presentation at the Synners meeting, May 27, 2009.

——. 2011. Proof theoretic background for linear grammar. MS., Ohio State University.

——, & E. ALLYN SMITH. to appear. A unified analysis of *the same*, phrasal comparatives and superlatives. In *Proceedings of SALT 2012*, volume ??, ??–??

REAPE, MIKE. 1996. Getting things in order. In *Discontinuous Constituency*, ed. by Harry Bunt & Arthur van Horck, volume 6 of *Natural Language Processing*, 209–253. Berlin, Germany and New York, NY, USA: Mouton de Gruyter. Published version of a Ms. from 1990.

SAG, IVAN, & RUI CHAVES. 2008. Left- and right-periphery ellipsis in coordinate and non-coordinate structures. MS., Stanford University and University at Buffalo, The State University of New York.

STEEDMAN, MARK. 1985. Dependency and coordination in the grammar of Dutch and English. *Language* 61.523–568.

YATABE, SHÛICHI. 2001. The syntax and semantics of left-node raising in Japanese. In *Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Dan Flickinger & Andreas Kathol, 325–344, Stanford. CSLI. http://cslipublications.stanford.edu/HPSG/.

YATABE, SHÛICHI. 2012. Comparison of the ellipsis-based theory of non-constituent coordination with its alternatives. In *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, Chungnam National University Daejeon*, ed. by Stefan Müller, 453–473.

ZAENEN, ANNIE, & LAURI KARTTUNEN. 1984. Morphological non-distinctiveness and coordination. In *Proceedings of the First Eastern States Conference on Linguistics*, ed. by Gloria Alvarez, Belinda Brodie, & Terry McCoy, 309–320.

A Coordination of unlikes with different subcategorization frames

In this appendix, we show that, by adopting the 'semantically potent' variant of the meet connective (in Bayer's (1996) terminology), examples such as (51) receives a straightforward analysis in the direct coordination analysis in CG, thereby refuting the claim occasionally raised in the literature by proponents of the LBE approach coordination that such examples undermine the CG analysis of NCC.

(51) John gave Mary a book and to Peter a record.

We assume that (51) is a variant of (52) which has undergone a surface-oriented reordering operation. For expository convenience, we provide the derivation for (52), and gloss over the details of the reordering operation.

(52) ?John gave Mary a book and a record to Peter.

The semantically potent variant of the meet connective assigns pairs of meanings as the denotations of the linguistic expressions that are assigned such categories. Thus, on this analysis, the different subcategorization frames of *give* can be compiled into one lexical entry in the following form:

(53) gave; $\langle \lambda x \lambda y \lambda z.\mathbf{give}(x)(y)(z), \lambda x \lambda y \lambda z.\mathbf{give}(y)(x)(z) \rangle$; (VP/PP/NP)∧(VP/NP/NP)

Note that, corresponding to the two subcategorization frames enocded in the syntactic category $\mathrm{VP/PP/NP}$ and $\mathrm{VP/NP/NP}$, we have two distinct semantic translations involving the same constant **give** but which take the first two arguments in different orders.

In actual derivations, one of these subcategorization frames is chosen via Meet Elimination:

(54)
$$\dfrac{\mathsf{john};\mathbf{j};\mathrm{NP} \quad \dfrac{\dfrac{\dfrac{\mathsf{gave};\ \langle \lambda x \lambda y \lambda z.\mathbf{give}(x)(y)(z), \lambda x \lambda y \lambda z.\mathbf{give}(y)(x)(z) \rangle;\ (\mathrm{VP/PP/NP}) \wedge (\mathrm{VP/NP/NP})}{\mathsf{gave};\ \lambda x \lambda y \lambda z.\mathbf{give}(y)(x)(z);\ \mathrm{VP/NP/NP}}\wedge\mathrm{E} \quad \mathsf{mary};\mathbf{m};\mathrm{NP}}{\mathsf{gave} \circ \mathsf{mary};\ \lambda y \lambda z.\mathbf{give}(y)(\mathbf{m})(z);\ \mathrm{VP/NP}}/\mathrm{E} \quad \mathsf{the} \circ \mathsf{book};\mathbf{b};\mathrm{NP}}{\mathsf{gave} \circ \mathsf{mary} \circ \mathsf{the} \circ \mathsf{book};\ \lambda z.\mathbf{give}(\mathbf{b})(\mathbf{m})(z);\ \mathrm{VP}}/\mathrm{E}}{\mathsf{john} \circ \mathsf{gave} \circ \mathsf{mary} \circ \mathsf{the} \circ \mathsf{book};\ \mathbf{give}(\mathbf{b})(\mathbf{m})(\mathbf{j});\ \mathrm{S}}\backslash\mathrm{E}$$

With these assumptions, the derivation for (52) is now straightforward. It just involves an interaction of the usual hypothetical reasoning analysis of NCC and the meet elimination analysis of the 'disambiguation' of two subcategorization frames assigned to *give* in (53). (Here, DTV abbreviates $\mathrm{VP/NP/NP}$ and PDTV abbreviates $\mathrm{VP/PP/NP}$; π_1 and π_2 are the first and second projection functions.)

(55)

$$\dfrac{\dfrac{\dfrac{\dfrac{[\varphi; f; \mathrm{PDTV} \wedge \mathrm{DTV}]^1}{\varphi; \pi_2(f); \mathrm{DTV}}\wedge\mathrm{E} \quad \mathsf{mary}; \mathbf{m}; \mathrm{NP}}{\varphi \circ \mathsf{mary}; \pi_2(f)(\mathbf{m}); \mathrm{VP/NP}}/\mathrm{E} \quad \mathsf{the} \circ \mathsf{book}; \mathbf{b}; \mathrm{NP}}{\varphi \circ \mathsf{mary} \circ \mathsf{the} \circ \mathsf{book}; \pi_2(f)(\mathbf{m})(\mathbf{b}); \mathrm{VP}}/\mathrm{E}}{\begin{array}{l}\mathsf{mary} \circ \mathsf{the} \circ \mathsf{book};\\ \lambda f.\pi_2(f)(\mathbf{m})(\mathbf{b});\\ (\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP}\end{array}}|\mathrm{I}^1 \quad \dfrac{\begin{array}{l}\mathsf{and};\\ \lambda g\lambda h.g \sqcap h;\\ (X\backslash X)/X\end{array} \quad \begin{array}{c}\vdots\ \ \vdots\\ \begin{array}{l}\mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda f.\pi_1(f)(\mathbf{r})(\mathbf{p});\\ (\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP}\end{array}\end{array}}{\begin{array}{l}\mathsf{and} \circ \mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda h.[\lambda f.\pi_1(f)(\mathbf{r})(\mathbf{p})] \sqcap h;\\ ((\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP})\backslash((\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP})\end{array}}/\mathrm{E}}{\begin{array}{l}\mathsf{mary} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda f.\pi_1(f)(\mathbf{r})(\mathbf{p}) \wedge \pi_2(f)(\mathbf{m})(\mathbf{b});\\ (\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP}\end{array}}\backslash\mathrm{E}$$

$$\dfrac{\begin{array}{l}\mathsf{gave};\\ \langle\lambda x\lambda y\lambda z.\mathbf{give}(x)(y)(z), \lambda x\lambda y\lambda z.\mathbf{give}(y)(x)(z)\rangle;\\ \mathrm{PDTV} \wedge \mathrm{DTV}\end{array} \quad \begin{array}{c}\vdots\ \ \vdots\\ \begin{array}{l}\mathsf{mary} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda f.\pi_1(f)(\mathbf{r})(\mathbf{p}) \wedge \pi_2(f)(\mathbf{m})(\mathbf{b});\\ (\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP}\end{array}\end{array}}{\begin{array}{l}\mathsf{gave} \circ \mathsf{mary} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda x.\mathbf{give}(\mathbf{r})(\mathbf{p})(x) \wedge \mathbf{give}(\mathbf{b})(\mathbf{m})(x);\\ \mathrm{VP}\end{array}}\backslash\mathrm{E}$$

Now, we should recall that, in defining the semantics of the meet connective, Bayer (1996) (section 7.1) opts for an impoverished, semantically nonpotent definition. The alleged motivation for this choice, according to Bayer (1996), comes from the fact that examples like the following (constituting a violation of Zaenen & Karttunen's (1984) well-known Anti-Pun Ordinance) can be derived once we admit the semantically potent variant of the meet connective and assign to *can* a syntactic category $(\mathrm{VP/NP}) \wedge (\mathrm{VP/VP[BASE]})$ with the corresponding semantic interpretation which pairs the main verb meaning and the auxiliary meaning as one entry.

(56) *I can tuna for a living and get a job if I want.

We take it that Bayer's (1996) argument here is somewhat misguided. In particular, Bayer (1996) seems to overlook the point that the availability of the semantically potent variant of the meet connective in the theory does not necessitate an analysis of the ambiguity of *can* in terms of it. The auxiliary *can* and the main verb *can* can just be entered in the lexicon as two separate entries, and then the overgeneration of (56) does not arise.

This of course begs the question of how to restrict the use of the semantically potent variant of the meet connective. In order to provide a complete answer to this question, we need to study examples of the sort exemplified by (52) more closely, but a conceptually plausible hypothesis which gives us a starting point is readily available: what distinguishes the ungrammatical cases like (56) and the grammatical cases like (52) seems to be that,

while the two uses of the same phonological form are totally unrelated in the former, the two subcategorization frames of *give* in the latter are clearly related semantically. By assuming that a semantically potent variant of the meet connective is invoked only in cases like the latter, we have a principled explanation for the Anti-Pun Ordinance while at the same time recognizing semantically potent meet. In other words, Anti-Pun Ordinance is a reflection of some substantive generalization governing the lexicon (whose exact nature we still don't understand), and not a consequence of a formal property of the underlying logic.

PERCEIVED FOREIGN ACCENT IN THREE VARIETIES OF NON-NATIVE ENGLISH*

Elizabeth A. McCullough
Ohio State University

Abstract

What aspects of the speech signal cause listeners to perceive a foreign accent? While many studies have explored this question for a single variety of non-native speech, few have simultaneously considered non-native speech from multiple native language backgrounds. In this perception study, American English-speaking listeners rated stop-vowel sequences extracted from English words produced by L1 American English, L1 Hindi, L1 Korean, and L1 Mandarin talkers on a continuous scale of degree of foreign accent. Stepwise linear regression models revealed that VOT, vowel quality, f0, and vowel duration contributed significantly to the ratings. Additionally, listeners rated productions by all varieties of non-native talkers as sounding foreign-accented to some degree, with those by L1 Hindi talkers as most foreign-accented, and those by L1 Mandarin talkers as more foreign-accented than those by L1 Korean talkers. The results suggest that several acoustic properties contribute substantially to the perception of foreign accent, at least for stop-vowel sequences, and that some varieties of non-native English sound more accented than others.

*I thank Mary Beckman, Cynthia Clopper, and Jeff Holliday for valuable input as this project developed, and audiences at the 161st meeting of the ASA and LabPhon13 for helpful comments on previous versions of this work.

Elizabeth A. McCullough. Perceived foreign accent in three varieties of non-native English. *OSUWPL* Volume 60, Spring 2013, 51-66.

1 Introduction

Native listeners of a language can recognize when someone is speaking that language with a foreign accent, even on the basis of very short samples of speech (Flege, 1984). However, they need not to do so accurately to motivate the study of foreign accent perception. Even when listeners misidentify a native talker as being foreign, or an L2 talker as being native, they are basing these judgments on some principled idea of what constitutes "foreign accent." Further, these judgments have repercussions, as sounding accented can have negative social consequences for the talker (Gluszek & Dovidio, 2010), and a listener viewing a talker as "other" may be biased against understanding the talker's speech (Rubin, 1992). The goals of the present investigation are to identify which acoustic properties contribute to the perception of foreign accent by native listeners of American English, and to explore such listeners' implicit views about the relative degrees of foreign accent in different varieties of non-native speech.

In many studies of perceived foreign accent, native listeners quantify the degree of foreign accent they hear in each production, and relationships between these ratings and talker-specific characteristics are examined. For instance, Oyama (1976) found a clear relationship between perceived foreign accentedness and age of arrival of Italian immigrants to the United States: the earlier an individual had immigrated, the weaker a foreign accent he was later judged to have. Flege, Munro, and MacKay (1995) conducted a similar investigation of Italian-accented English, and found that age of learning accounted for 59% of the variance in perceived foreign accent ratings, with earlier learners sounding more native. Such studies, however, do not address the question of which characteristics directly influence native listeners in their assignments of perceived foreign accent ratings. Listeners have no knowledge of an individual talker's language history, and must be attending to properties of the acoustic signal.

What acoustic properties might contribute to the perception of foreign accent? Traditional accounts of L2 acquisition (e.g., Lado, 1957) refer to cross-language phonological differences and the role of L1 "interference" in production of the L2. This suggests that it might be possible to measure and compare acoustic properties that are likely to differ between a talker's L1 and L2, and correlate the degree of perceived foreign accent with the degree of difference on these specific measures.

While this approach is not new, investigations of the signal have generally focused on single acoustic properties. For instance, VOT has been found to contribute to the perception of foreign accent for voiceless stops in L2 Spanish productions by L1 English speakers (Gonzalez-Bueno, 1997), and in L2 English productions by L1 Brazilian Portuguese speakers (Major, 1987) and L1 Japanese speakers (Riney & Takagi, 1999). Pronunciations of liquids and vowels seemed to influence the perception of foreign accent in L2 English productions by L1 Japanese speakers in Riney, Takagi, and Inutsuka (2005), although this was determined by a phonetically trained listener's auditory analysis rather than any acoustic measurements. Munro and Derwing (2001) found that speaking rate influenced the perception of foreign accent in the speech of L2 English talkers from 12 different L1s.

In addition, such investigations, with the notable exception of Munro and Derwing (2001), have nearly exclusively focused on single varieties of non-native speech. Thus, it is not clear that these studies investigated "foreign accent" generally as opposed to more

specific scales such as "Brazilian Portuguese accent" or "Japanese accent." If the task demanded only comparisons between native talkers and L2 talkers from a single L1 background, listeners might have implemented the more specific scale regardless of the instructions given.

To focus on "foreign accent" in general as opposed to some particular accent, the present study, like Munro and Derwing (2001), uses L2 English productions by talkers of multiple L1s. In addition, multiple acoustic properties are measured and evaluated in relation to listeners' ratings of the degree of foreign accent in each of these productions. The analysis below will identify acoustic properties that seem to influence foreign accent perception, and determine whether some L1 backgrounds give rise to a stronger percept of foreign accent in L2 English than do others.

2 Methods

2.1 Acoustic stimuli

The acoustic stimuli in this experiment were extracted from recordings in the Buckeye GTA Corpus (Hardman, 2010). This corpus includes productions of 64 of the Bamford-Kowal-Bench sentences revised for American English (Bamford & Wilson, 1979) by 24 L1 American English talkers, 19 L1 Hindi talkers, 20 L1 Korean talkers, and 20 L1 Mandarin talkers. All non-native talkers were of a reasonably high English proficiency level, in that they were certified as a Graduate Teaching Associate at The Ohio State University (by a score of 230/300 on the SPEAK test or an "unconditional pass" on the university's Mock Teaching Test) or had scored at least 26/30 on the speaking section of the TOEFL iBT. For this study, 4 female talkers from each of the 4 L1 groups were used, for a total of 16 talkers. Including 4 talkers from each L1 ensured that there was variation within each L1 group, such that any effects of L1 would not be confounded with individually idiosyncratic pronunciations. Although English is commonly spoken in India, none of the 4 L1 Hindi talkers chosen identified English as a native language.

While many previous studies have used sentences as audio stimuli in perceived foreign accent rating tasks, an exhaustive acoustic investigation of sentences would be a daunting undertaking. Units as large as sentences have a relatively large number of possible segmental and suprasegmental cues, many of which might be expected to influence ratings of perceived foreign accent. In order to limit the number of potentially relevant acoustic cues, stop-vowel sequences were chosen as the acoustic stimuli for the present study. Stop-vowel sequences are even smaller than the words rated in Gonzalez-Bueno (1997) and Major (1987), and as such have a comparatively small number of possible cues to explore, although listeners are still able to perceive foreign accent in units this size (Flege, 1984).

Sequences containing stops are particularly interesting given the language backgrounds included in the Buckeye GTA Corpus. Stops in some of these languages differ from those in American English phonologically: at each place of articulation, Korean distinguishes 3 stops (Lee, 1999) and Hindi, 4 (Ohala, 1999), while American English and Mandarin have only 2 (Ladefoged, 1999; Lee & Zee, 2003). Additionally, the stop contrasts in Hindi, Korean, and Mandarin all differ acoustically from the American English contrast. Such differences might be reflected to some degree in the English productions of these non-native talkers.

The stop-vowel productions used as stimuli were extracted from utterance-medial, word-initial contexts in the Buckeye GTA Corpus recordings. Three stop-vowel sequences were chosen for each of the 6 American English stops, for a total of 18 stop-vowel sequences. The limited number of word types in the corpus did not allow for control of the vowel, nor for the exclusion of stop-vowel sequences that were themselves lexical items (/pi/ and /tu/). All 18 stop-vowel sequences and the words they were extracted from are included in Table 1 in the Appendix.

Each of the 16 talkers produced each of the 18 stop-vowel sequences, for a total of 288 audio stimuli. The mean intensity of all audio stimuli was normalized. No fillers were used.

2.2 Participants

28 monolingual American English-speaking listeners participated in the rating study for linguistics course credit.

2.3 Task

The audio stimuli were presented through headphones at individual computer stations. Listeners were asked to rate the degree of foreign accent in each audio stimulus by sliding a bar along a continuous Visual Analog Scale (VAS). The endpoints of the VAS were labeled "no foreign accent" and "strong foreign accent," as shown in Figure 1. Listeners were encouraged to use the scale continuously rather than categorically. VAS was chosen because listeners have fairly sensitive responses to foreign accent rating tasks, as confirmed by the continuous rating scale used by Flege et al. (1995), which might not be sufficiently captured by a Likert scale with a fixed number of intervals. In addition, a continuous rating scale allowed for the possibility of better correlation with the continuous acoustic cues discussed below.

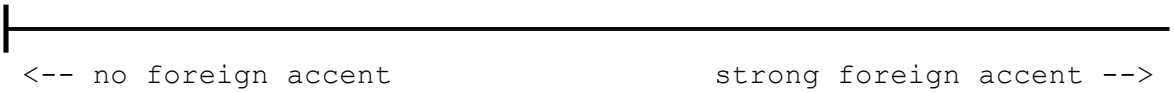

Figure 1. VAS rating line

Just before each stimulus played, the word it had been extracted from was displayed orthographically on the computer screen so that listeners had some perceptual target for the upcoming production. The audio stimuli were blocked by stop-vowel sequence; that is, all 16 productions of the same stop-vowel sequence were played consecutively. Listeners were permitted to take breaks between each block. Block order was randomized, as was stimulus order within each block. A practice block with an additional stop-vowel sequence not used as a test stimulus (/be□/ from baby) was administered before the test blocks so that listeners could experience the types of variation in the stimuli and practice using the rating scale.

2.4 Acoustic properties

Four acoustic properties were considered in the analysis: VOT, vowel quality, vowel duration, and f0 (fundamental frequency). VOT and vowel quality (midpoint F1 and F2)

were chosen as fairly standard measures that have been previously shown to be related to perceived foreign accent ratings (Gonzalez-Bueno, 1997; Major, 1987; Riney & Takagi, 1999; Riney et al., 2005). Vowel duration was included in response to remarks from listeners that some stimuli were too short to rate properly. Similarly, f0 was included in response to remarks from listeners that differences in pitch made the task difficult.

Speaking rate, although found to play a role in foreign accent perception by Munro and Derwing (2001), was not investigated here. Because a temporal measure of each of the two segments in each stop-vowel stimulus was already included, speaking rate would have been largely redundant.

2.5 Analysis

For the production measures, VOT and vowel intervals were marked by the author in Praat (Boersma & Weenink, 2012), and their durations were extracted automatically. F1 and F2 values at the vowel midpoint and mean f0 values over the vowel were extracted automatically using a Praat script that displayed a spectrogram with overlaid formant and pitch tracks, such that the author could check token-by-token for tracking errors. In the rare case of such an error, the author re-measured the token in question with modified settings.

For the perception measure, each listener rating was recorded by the experiment presentation software as a numerical value representing the position on the rating line. In total, 8064 ratings were assigned (16 talkers x 18 stop-vowel sequences x 28 raters).

3 Predictions

While variation in both acoustic measurements and perceived foreign accent ratings can of course be expected in productions by different talkers from the same L1 background and even in different productions by the same talker, this study focuses mainly on variation across L1 backgrounds. Actual language background is expected to align to some degree with perceived foreign accent in a talker's speech, in that productions from native talkers are likely to be judged as exhibiting little to no foreign accent, and productions by non-native talkers are likely to be judged as exhibiting some degree of foreign accent. In this section, the effect of L1 background on each acoustic property is examined for the stimuli used in the present study, and subsequent predictions are made. If there are differences between native and non-native productions for a particular acoustic property, then listeners might use that property to cue foreign accent. Additionally, if a particular variety of non-native speech differs from native speech on multiple acoustic properties, then it might be judged as strongly foreign-accented. As explained further in Section 4.1, stimuli with voiced versus voiceless stop targets will be analyzed separately to facilitate comparison with previous studies; thus, these acoustic measurements are also presented separately for the two groups of stimuli.

3.1 Differences by L1 background

Figure 2 shows VOT values for the productions by talkers of each variety of English. Separate one-way ANOVAs with talkers as subjects and L1 as a between-subjects factor revealed significant effects of L1 on VOT for voiced ($F(3,12) = 6.50$; $p < 0.01$) and voiceless ($F(3,12) = 12.80$; $p < 0.001$) stop targets. Separate Tukey post-hoc tests for

voiced and voiceless stop targets showed that these effects were driven by the differences between L1 Hindi talkers' values and all others ($p < 0.05$); no other pairwise comparisons were significant. Indeed, the most striking details of Figure 2 are the VOT ranges for L1 Hindi talkers' productions. Unlike the voiced stop productions by talkers from other L1 backgrounds, which exhibited short lag voicing, a substantial portion (56%) of the L1 Hindi talkers' voiced stop productions were actually prevoiced. Additionally, L1 Hindi talkers' voiceless stop productions had much shorter VOT values than those for talkers of all other L1s.

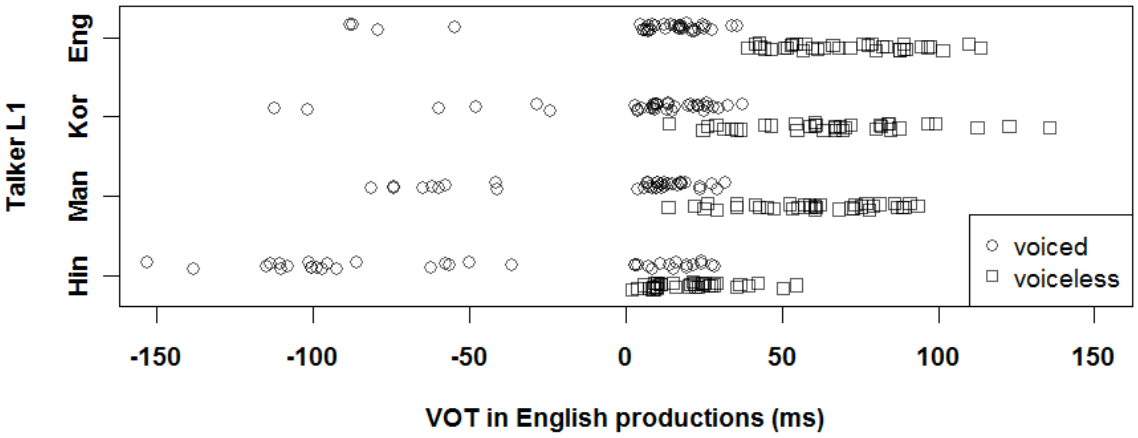

Figure 2. VOT values by target voicing and talker L1

The use of multiple varieties of non-native English in this study presents a problem for linear regression using raw values: productions by native talkers can be expected to have values near one extreme of the range for perceived foreign accent ratings ("no foreign accent"), but the same assumption cannot necessarily be made about the distribution of their acoustic measurements. That is, it would be possible for productions by talkers from one non-native background to have values lower than those observed in productions by native talkers, while productions by talkers from a different non-native background might have values higher than those observed in productions by native talkers. This problem is avoided here by using values that represent the difference between a production and the comparable native productions on some acoustic property. Specifically, all acoustic cues considered below are parameterized as the absolute difference from the American English mean, calculated separately for each of the 18 stop-vowel sequences. For instance, rather than examining the raw VOT value for some L1 Korean talker's production of /pi/ from the word people, the value considered is the absolute difference between the VOT value of this production of /pi/ and the mean VOT of the 4 L1 American English talkers' productions of /pi/. Difference values for productions by L1 American English talkers are also calculated, and should generally be small. Figure 3 shows VOT differences for the productions by each talker. Talkers within each L1 group are presented in a consistent order across all plots.

Separate one-way ANOVAs with talkers as subjects and L1 as a between-subjects factor revealed significant effects of L1 on VOT difference for voiced ($F(3,12) = 7.45$; $p < 0.01$) and voiceless ($F(3,12) = 38.38$; $p < 0.001$) stop targets. Again, this result was due to the values for productions by L1 Hindi talkers being different from those for productions by talkers from all other L1 backgrounds, as shown by separate Tukey post-hoc tests for voiced and voiceless stop targets ($p < 0.05$). In the case of VOT, the effect

of L1 background is the same regardless of whether raw values or values that denote the difference from native production means are used.

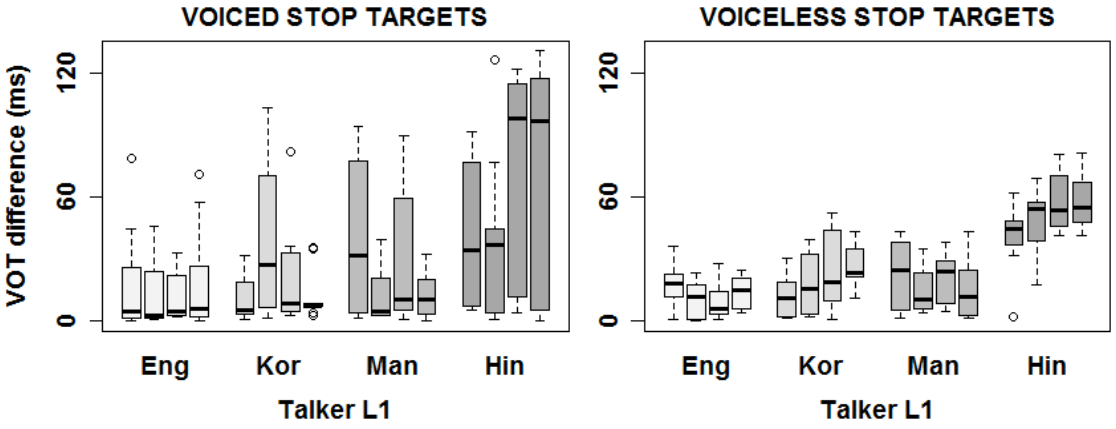

Figure 3. VOT differences by target voicing, talker L1, and talker identity

Vowel quality is defined in terms of two dimensions, F1 and F2. To reduce this to a single dimension, vowel quality difference is quantified as the Euclidean distance in F1/F2 space from the position of a single production to the mean position for the productions of that stop-vowel sequence by the L1 American English talkers. Figure 4 shows vowel quality differences for the productions by each talker. A one-way ANOVA with talkers as subjects and L1 as a between-subjects factor showed a significant effect of L1 on vowel quality difference for stimuli with voiceless stop targets ($F(3,12) = 8.04$; $p < 0.01$); the effect was not significant for stimuli with voiced stop targets. A Tukey post-hoc test for stimuli with voiceless stop targets revealed that vowel quality difference in productions by L1 Korean and L1 Mandarin talkers was different from that in productions by L1 American English talkers ($p < 0.05$); difference values were greater for productions by the non-native talkers. The comparison between difference values for L1 Hindi talkers' productions and L1 American English talkers' productions, in the same direction, approached significance ($p = 0.05$).

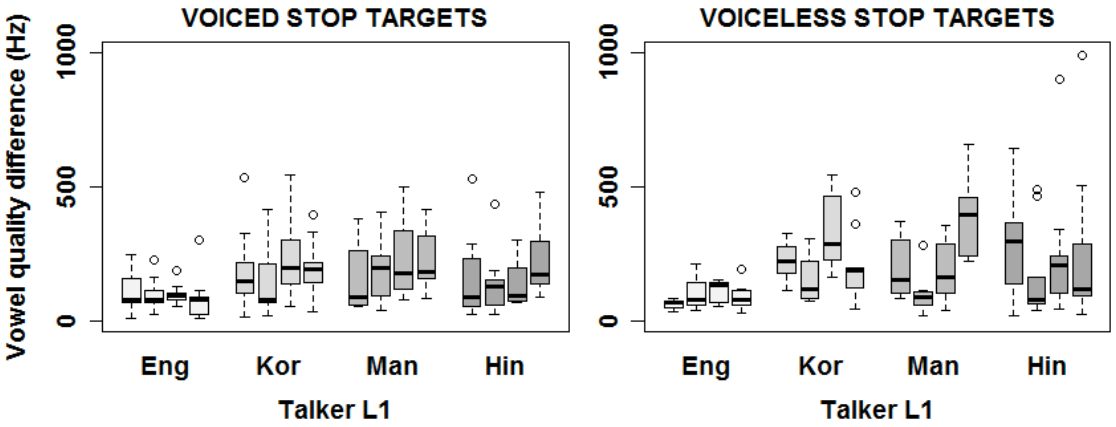

Figure 4. Vowel quality differences by target voicing, talker L1, and talker identity

Figure 5 shows vowel duration differences for the productions by each talker. A one-way ANOVA with talkers as subjects and L1 as a between-subjects factor revealed a

significant effect of L1 on vowel duration difference for stimuli with voiced stop targets ($F(3,12) = 10.12$; $p < 0.01$); the effect was not significant for stimuli with voiceless stop targets. A Tukey post-hoc test for stimuli with voiced stop targets showed that vowel duration difference in productions by L1 Korean and L1 Mandarin talkers was different from that in productions by L1 American English talkers ($p < 0.05$); again, difference values were greater for productions by the non-native talkers.

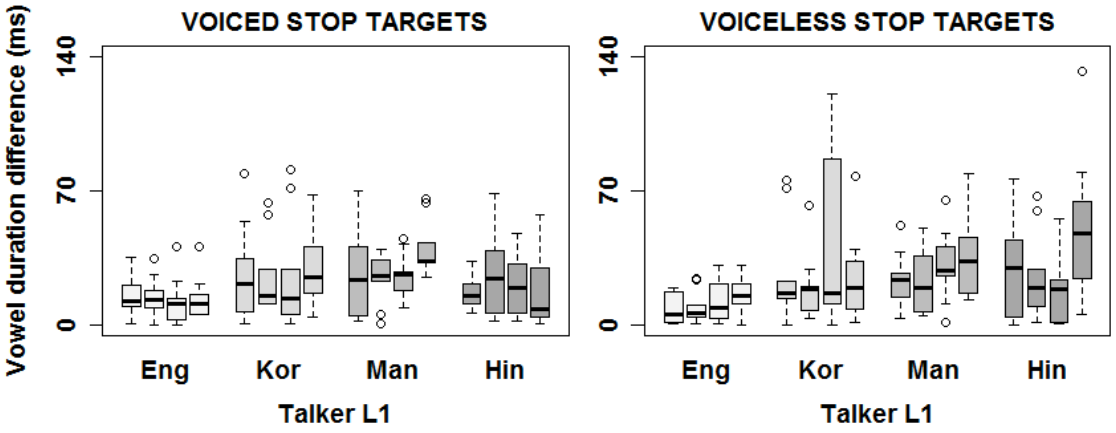

Figure 5. Vowel duration differences by target voicing, talker L1, and talker identity

Figure 6 shows f0 differences for the productions by each talker. Separate one-way ANOVAs with talkers as subjects and L1 as a between-subjects factor did not reveal a significant effect of L1 on f0 difference for either set of stimuli.

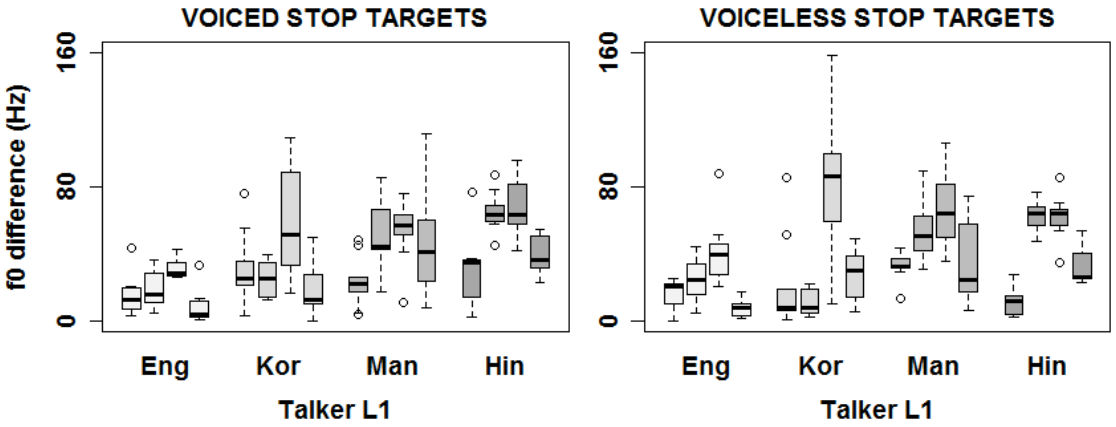

Figure 6. f0 differences by target voicing, talker L1, and talker identity

3.2 Acoustic cues

As differences were observed between native and non-native talkers in VOT, vowel quality, and vowel duration, these acoustic properties might play a role in foreign accent perception. In support of this prediction, the relationship of VOT to foreign accent perception has been demonstrated previously by acoustic measurements (Gonzalez-Bueno, 1997; Major, 1987; Riney & Takagi, 1999), and vowel quality's role has been suggested previously by auditory analysis (Riney et al., 2005). In contrast, f0 did not

reliably differ by L1 background and may be unlikely to influence the perception of foreign accent.

3.3 Overall rating patterns

Productions by L1 Hindi talkers differed from native talkers' productions in VOT and marginally in vowel quality, while productions by L1 Korean and L1 Mandarin talkers differed from native talkers' productions in vowel quality and vowel duration. Due to these acoustic differences, it is expected that productions from all non-native backgrounds should sound accented to some degree. As VOT has repeatedly been shown to correlate with perceived foreign accent ratings, and as productions from L1 Hindi talkers are the only ones that came close to showing effects for properties of both consonants and vowels, it is possible that productions from L1 Hindi talkers might sound more foreign-accented than productions by other non-native talkers. This would align well with feedback from listeners, who often commented that the "Indian accent" was the most obvious.

4 Results

In this section, the possible acoustic cues to foreign accent perception are evaluated, and then the relative strength of foreign accent in different varieties of non-native English is explored.

4.1 Acoustic cues

The regression models detailed in this section test the relationship between perceived foreign accent ratings and acoustic cues directly, and the stepwise models evaluate the contributions of all measured acoustic cues simultaneously. A significant positive relationship between any acoustic cue and perceived foreign accent ratings means that as values of that acoustic cue deviate more from the L1 American English talkers' mean, listeners perceive a higher degree of foreign accent.

The dependent variable is the mean perceived foreign accent rating for each of the 288 stimuli. Thus, while the predictions above were based on differences in acoustic properties across L1 groups, this portion of the analysis should also account for differences in ratings among talkers in the same L1 group, as well as differences in ratings among the 18 productions by a single talker. To the extent that productions varied in some significant acoustic cue, in terms of deviation from the native mean, the ratings for those productions should be different, regardless of the identity of the talker.

Previous studies have found clear relationships between VOT and ratings of perceived foreign accent (Gonzalez-Bueno, 1997; Major, 1987; Riney & Takagi, 1999). Notably, however, these studies only explored this relationship with voiceless stop targets. Thus, the results below are presented separately for voiced and voiceless stop targets, to make comparison with previous findings possible.

A stepwise linear regression model of the data for stimuli with voiced stop targets only identified vowel quality, VOT, f0, and vowel duration, in that order, as significant cues. Four linear regression models were built, such that the first included only the most significant cue, the second included the two most significant cues, etc. Full statistical

details are given in Table 2 in the Appendix. For the present discussion, it is important to highlight that the model with vowel quality accounted for 18% of the variance in the perceived foreign accent ratings. With the addition of VOT, this rose to 36% (18% improvement); with the addition of f0, this rose to 42% (6% improvement); and with the addition of vowel duration, this rose to 44% (2% improvement). Thus, the best model accounted for less than half of the variance in ratings, and included an acoustic property, f0, that was not predicted to play a role.

Similar models were created for stimuli with voiceless stop targets only; full statistical details are given in Table 3 in the Appendix. In the first model, VOT alone explained 40% of the variance in the ratings of perceived foreign accent—nearly as much as all four acoustic cues combined in the model for voiced targets. The addition of vowel quality brought this value to 45% (5% improvement), and the model that also included f0 accounted for 48% of the variance (3% improvement). Vowel duration, although significant in the stepwise model, did not contribute significantly to the explicitly specified regression model.

4.2 Overall rating patterns

Perceived foreign accent ratings by talker are shown for stimuli with voiced and voiceless stop targets separately in Figure 7. Separate one-way repeated measures ANOVAs with listeners as subjects and L1 as a within-subjects factor revealed significant effects of L1 on perceived foreign accent ratings for stimuli with voiced ($F(3,104) = 8.93$; $p < 0.001$) and voiceless ($F(3,104) = 10.29$; $p < 0.001$) stop targets. Post-hoc Bonferroni-corrected paired t-tests showed differences between all pairwise comparisons for voiced targets ($p < 0.00017$), as well as for voiceless targets ($p < 0.00017$). While there is quite a range of ratings within each L1 group, general patterns do emerge. As predicted, productions by L1 American English talkers were rated as having very little foreign accent, and productions by non-native talkers were rated as having greater degrees of foreign accent. In addition, productions by L1 Hindi talkers were rated as having the highest degree of foreign accent. While there was no basis for predicting a difference between perceived foreign accent ratings for productions by L1 Korean talkers as compared to productions by L1 Mandarin talkers, the latter received higher ratings, as shown in Figure 7.

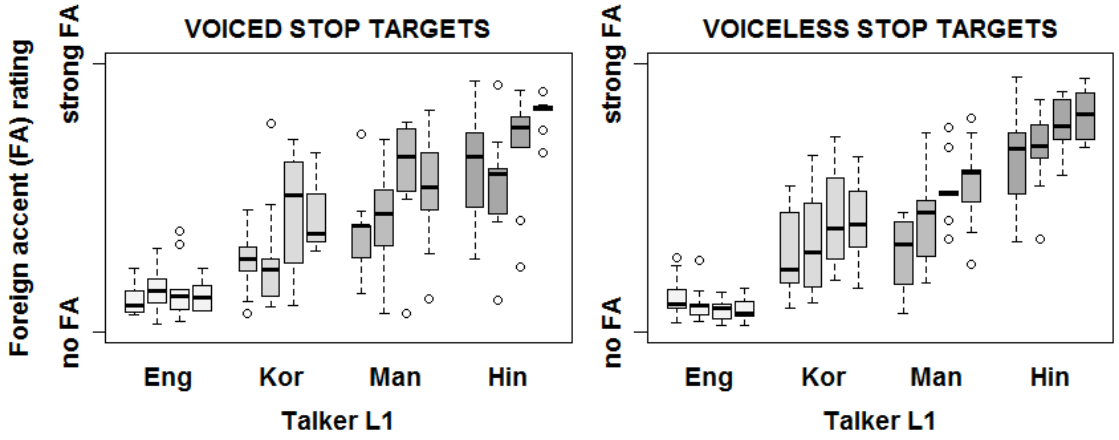

Figure 7. Ratings of perceived foreign accent by target voicing, talker L1, and talker identity

5 Discussion

The first goal of this research was to identify potential acoustic cues to perceived foreign accent. Acoustic properties were measured in very short stop-vowel sequences extracted from sentences in English. The same sequences were used as stimuli in a rating task, and the perceived foreign accent ratings were shown to be correlated with VOT, vowel quality, f0, and vowel duration for at least some of the stimuli. For stimuli with voiced stop targets, the best model included all four acoustic properties and accounted for 44% of the variance in perceived foreign accent ratings. For stimuli with voiceless stop targets, VOT, vowel quality, and f0 accounted for 48% of the variance in perceived foreign accent ratings, with VOT playing a substantial role.

The present study confirmed previous findings that VOT and vowel quality seem to be involved in the perception of foreign accent, suggested that f0 and vowel duration may also play minor roles, and showed that the relative importance of these cues differs somewhat for stimuli containing voiced stops targets as opposed to voiceless stop targets. This last fact is particularly notable given the frequent use of sentence-length stimuli, which contain many segments that may or may not be controlled for voicing, in studies of foreign accent. It remains to be seen whether such differences in relative importance persist in the perception of longer stimuli, or whether segmental cues are washed out by more global prosodic properties.

Although VOT was a significant predictor regardless of the subset of data analyzed, it clearly played a larger role in ratings of stimuli with voiceless stop targets than in ratings of stimuli with voiced stop targets. Figure 2 showed that a number of the voiced stop productions, especially by L1 Hindi talkers, were prevoiced, while voiceless stop productions were characterized by lag voicing. The difference in the contribution of VOT might have resulted from the listener population in this study; American English-speaking listeners are likely to be quite sensitive to lag voicing differences, as most word-initial productions by native speakers of American English are in this range. However, this effect might equally have been a general consequence of perception, as even listeners of languages with prevoiced stops seem to be more sensitive to the presence or absence of prevoicing than to the amount of prevoicing (van Alphen & McQueen, 2006). Thus, a convincing linear relationship between VOT and perceived foreign accent ratings might be unlikely for any stimuli with voiced stop targets. If so, a linear regression analysis would not be expected to perform well on this portion of the data. As previous investigations have generally focused only on voiceless stop targets, this complication has not been widely recognized.

While all four acoustic cues measured were found to contribute significantly to the model for at least one of the two target voicing categories, collectively they accounted for around half of the variance in perceived foreign accent ratings, leaving half of the variance unexplained. Stop-vowel stimuli were chosen so as to limit the number of acoustic cues listeners were able to attend to in performing the task, but it seems that listeners used more than the four acoustic cues measured here. As an individual's L1 is expected to influence L2 production, additional acoustic properties to consider in future analyses might be those that distinguish stops in the non-native talkers' L1s. For instance, voice quality and noise offset time are often invoked in phonetic descriptions of Hindi stops (Davis, 1994), and in Korean, voice quality and f0 at the onset of the following vowel are relevant to the 3-way stop distinction (Kong, Beckman, & Edwards,

2011). Although such cues are not needed for phonemic distinctions in American English, this does not mean that they were not perceived by American English-speaking listeners. If detectable in the stimuli, such cues might clearly identify certain productions as foreign and augment their ratings of perceived foreign accent.

The second goal of this research was to investigate whether some varieties of non-native English sound generally more accented than others. The large ranges of ratings for productions by talkers of each L1 indicate that listeners were not simply categorizing the productions into four discrete piles on the scale. Acoustic measurements differ among productions by talkers of different L1s, but also among productions by talkers from the same L1 and even among productions by the same talker. These multiple levels of variation might explain why f0 turned out to contribute to the regression models for both voiced and voiceless stop targets despite there being no effect of talker L1 on its difference values. Nonetheless, it seems clear that not all L1 backgrounds are equal when it comes to the perception of foreign accent. While productions by all non-native talkers were judged as sounding foreign-accented to some degree, those by L1 Hindi talkers were most accented, while those by L1 Mandarin talkers were more accented than those by L1 Korean talkers. These results suggest that listeners have some quantifiable concept of "foreign accent" more abstract than the "Brazilian Portuguese accent" or "Japanese accent" concepts that might have been explored in earlier studies, though the relationship between the abstract "foreign accent" and the various more specific incarnations is yet unclear.

These two sets of results also can be related to each other. For instance, the acoustic analysis showed that productions by L1 Hindi talkers differed significantly from productions by native talkers in their VOT values. Thus, there is a clear relationship between the importance of VOT as an acoustic cue and the high perceived foreign accent ratings for productions by L1 Hindi talkers. The directionality, however, is uncertain. Are productions by L1 Hindi talkers rated as most foreign-accented because they differ in VOT, or does VOT matter because English produced by L1 Hindi talkers is thought to sound highly foreign-accented? When asked if they could identify the accents present in the stimuli, 19 of 28 listeners identified "Indian" (15), "Hindi" (3), or "South Asian" (1), indicating that this variety of speech may have been easily identifiable. However, 18 of 28 listeners identified "Chinese" (9), "Mandarin" (1), "Asian" (7), or "East Asian" (1), while this variety of speech was rated as sounding significantly less foreign-accented. Only 2 of 28 listeners specifically identified "Korean," although some of the "Asian" and "East Asian" responses may have been at least partly in response to the productions by L1 Korean talkers. Further research is needed to investigate the role that stereotypes and familiarity with particular non-native varieties might play in the perception of foreign accent.

The results of this study are particularly interesting in light of the sociolinguistic situation of English in India, where "it is the 'associate official' language of the country and it also serves as a link language between the educated" (Gargesh, 2004, p. 992). Although they did not claim English as a native language, the L1 Hindi talkers in this study reported that they first began studying English between ages 3 and 4. In contrast, the L1 Korean and L1 Mandarin talkers in this study reported that they did not begin to learn English until at least age 11. Compared to the other non-native groups, the L1 Hindi talkers' experience with English started earlier and lasted longer. Their target variety of English was Indian rather than American, however, and the ratings from this

experiment suggest that this variety can sound quite foreign-accented to American English-speaking listeners. Thus, judgments about foreign accent are critically shaped by the listener's ideas about the world.

A listener's impressions regarding foreign accent develop over the course of day-to-day language use. However, most communication requires units much more complex than the stop-vowel sequences used as stimuli in this study, and the cues discovered here would not necessarily play the same roles in ratings of larger units like words or sentences. As longer stimuli would exhibit a much wider variety of potential acoustic cues, a considerably more complicated relationship between perceived foreign accent ratings and acoustic cues would seem inevitable. Nonetheless, the present work has taken the initial steps in addressing the question of what aspect of the speech signal causes listeners to perceive a foreign accent.

References

Bamford, John, and Ian Wilson. 1979. Methodological considerations and practical aspects of the BKB sentence lists. In *Speech-hearing tests and the spoken language of hearing-impaired children*, ed. by John Bench and John Bamford, 148-187. London: Academic Press.

Boersma, Paul, and David Weenink. 2012. *Praat: Doing phonetics by computer* (Version 5.3) [Computer program]. <http://www.praat.org/>.

Davis, Katharine. 1994. Stop voicing in Hindi. *Journal of Phonetics* 22.177-193.

Flege, James Emil. 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76.692-707.

Flege, James Emil, Murray Munro, and Ian MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97.3125-3134.

Gargesh, Ravinder. 2004. Indian English: Phonology. In *A Handbook of varieties of English*, ed. by Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie, and Clive Upton, 992-1002. Berlin: Mouton de Gruyter.

Gluszek, Agata, and John F. Dovidio. 2010. The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review* 14(2).214-237.

Gonzalez-Bueno, Manuela. 1997. Voice onset time in the perception of foreign accent by native listeners of Spanish. *International Review of Applied Linguistics in Language Teaching* 35(4).251-262.

Hardman, Jocelyn B. 2010. The intelligibility of Chinese-accented English to international and American students at a US university. Columbus, OH: OSU dissertation.

Kong, Eun Jong, Mary E. Beckman, and Jan Edwards. 2011. Why are Korean tense stops acquired so early: The role of acoustic properties. *Journal of Phonetics* 39(2).196-211.

Ladefoged, Peter. 1999. American English. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, 41-44. Cambridge: Cambridge University Press.

Lado, Robert. 1957. *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press: Ann Arbor.

Lee, Hyun Bok. 1999. Korean. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, 120-123. Cambridge: Cambridge University Press.

Lee, Wai-Sum, and Zee, Eric. 2003. Illustrations of the IPA: Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1).109-112.

Major, Roy C. 1987. English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics* 15.197-202.

Munro, Murray J., & Tracey M. Derwing. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech. *Studies in Second Language Acquisition* 23.451-468.

Ohala, Manjari. 1999. Hindi. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, 100-103. Cambridge: Cambridge University Press.

Oyama, Susan. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5.261-285.

Riney, Timothy J., and Naoyuki Takagi. 1999. Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning* 49(2).275-302.

Riney, Timothy J., Naoyuki Takagi, and Kumiko Inutsuka. 2005. Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly* 39(3).441-466.

Rubin, Donald L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33(4).511-531.

van Alphen, Petra M., and James M. McQueen. 2006. The effect of Voice Onset Time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance* 32(1).178-196.

Appendix

Table 1. Stop-vowel sequences used as audio stimuli

Sequence	*Word*
/bæ/	back
/bʌ/	buckets
/bɑ/	boxes
/dɪ/	dish
/dɪ/	dinner
/dɔ/	dog
/gɚ/	girl
/gʊ/	good
/gɑ/	got
/pi/	people
/pɪ/	picture
/pæ/	packed
/teɪ/	table
/tu/	two
/tɑʊ/	towel
/kɪ/	kitchen
/keɪ/	came
/koʊ/	coat

Table 2. Regression models (stimuli with voiced stop targets)

Variable	*Value*	*Statistical test*
Model 1	$R^2 = 0.18$	$F(1,142) = 32.67$, $p < 0.001$
Vowel quality	$m_1 = 0.09$	$t(142) = 5.72$, $p < 0.001$
Model 2	$R^2 = 0.36$	$F(2,141) = 41.06$, $p < 0.001$
Vowel quality	$m_1 = 0.10$	$t(141) = 6.72$, $p < 0.001$
VOT	$m_2 = 0.32$	$t(141) = 6.36$, $p < 0.001$
Model 3	$R^2 = 0.42$	$F(3,140) = 35.93$, $p < 0.001$
Vowel quality	$m_1 = 0.09$	$t(140) = 6.04$, $p < 0.001$
VOT	$m_2 = 0.28$	$t(140) = 5.78$, $p < 0.001$
f0	$m_3 = 0.28$	$t(140) = 4.07$, $p < 0.001$
Model 4	$R^2 = 0.44$	$F(4,139) = 28.91$, $p < 0.001$
Vowel quality	$m_1 = 0.08$	$t(139) = 5.64$, $p < 0.001$
VOT	$m_2 = 0.28$	$t(139) = 5.90$, $p < 0.001$
f0	$m_3 = 0.27$	$t(139) = 3.99$, $p < 0.001$
Vowel duration	$m_4 = 0.20$	$t(139) = 2.21$, $p < 0.05$

Table 3. Regression models (stimuli with voiceless stop targets)

Variable	*Value*	*Statistical test*
Model 1	$R^2 = 0.40$	$F(1,142) = 96.21$, $p < 0.001$
VOT	$m_1 = 0.85$	$t(142) = 9.81$, $p < 0.001$
Model 2	$R^2 = 0.45$	$F(2,141) = 60.02$, $p < 0.001$
VOT	$m_1 = 0.76$	$t(141) = 8.82$, $p < 0.001$
Vowel quality	$m_2 = 0.04$	$t(141) = 3.82$, $p < 0.001$
Model 3	$R^2 = 0.48$	$F(3,140) = 44.91$, $p < 0.001$
VOT	$m_1 = 0.73$	$t(140) = 8.69$, $p < 0.001$
Vowel quality	$m_2 = 0.03$	$t(140) = 3.35$, $p < 0.01$
f0	$m_3 = 0.17$	$t(140) = 2.90$, $p < 0.01$
Model 4	$R^2 = 0.48$	$F(4,139) = 34.57$, $p < 0.001$
VOT	$m_1 = 0.71$	$t(139) = 8.25$, $p < 0.001$
Vowel quality	$m_2 = 0.03$	$t(139) = 2.66$, $p < 0.01$
f0	$m_3 = 0.17$	$t(139) = 2.98$, $p < 0.01$
Vowel duration	$m_4 = 0.11$	$t(139) = 1.51$, n.s.

AN INTRODUCTION TO RANDOM PROCESSES FOR THE SPECTRAL ANALYSIS OF SPEECH DATA

Patrick F. Reidy
Ohio State University

Abstract

Spectral analysis of acoustic data is a common analytical technique with which phoneticians have ample practical experience. The primary goal of this paper is to introduce to the phonetician, whose primary interest is the analysis of linguistic data, a portion of the theory of random processes and the estimation of their spectra, knowledge of which bears directly on the choices made in the process of analyzing time series data, such as an acoustic waveform. The paper begins by motivating the use of random processes as a model for acoustic speech data, and then introduce the spectral representation (or, spectrum) of a random process, taking care to relate this notion of spectrum to one that is more familiar to phoneticians and speech scientists. A final section presents two methods for estimating the values of the spectrum of a random process. Specifically, it compares the commonly-used (windowed) periodogram to the multitaper spectrum, and it is shown that the latter has many beneficial theoretical properties over the former.

Patrick F. Reidy. An introduction to random processes for the spectral analysis of speech data. *OSUWPL* Volume 60, Spring 2013, pp. 67-116.

1 Introduction

This paper discusses some of the statistical methods involved in the spectral analysis of speech data. Specifically, its aim is to introduce phoneticians to random processes, the class of mathematical object used to model speech data; their spectral representation; and some of the methods for estimating the values of a random process's spectrum.

In order to appreciate the place that random processes hold in a spectral analysis of speech data, we first consider a concrete example of such an analysis. Suppose that a researcher wishes to investigate the spectral properties of the English voiceless sibilant /s/. The first step in this investigation is to collect data by recording several tokens of /s/ from multiple English speakers. We refer to this type of data as *speech data*, measurements of the air pressure fluctuations caused by a particular speech sound wave, as sensed by a microphone. In practice, these measurements are typically stored by a digital recording device as a sequence of numbers, where each number represents the instantaneous air pressure at a given time.

So, the actual physical sound wave generated during speech production and the experimenter's record of that sound wave differ in basic ways. Whereas, the physical sound wave causes continuous air pressure fluctuations over a continuous interval of time, the record of the sound wave has been both sampled and quantized, which results in discrete air pressure fluctuations that occur over a discrete time interval. Because the researcher has access to only the record of the sound wave and because our focus is the analysis of speech data, we choose to represent a sound wave and its waveform as a numeric sequence.

Figure 1 shows the waveform of a token of /s/ that might be recorded by the researcher. The values of this waveform appear to vary randomly from one sample to the next. This random variation is expected in the waveform of /s/ because its noise source is generated by turbulent airflow, which by definition involves random air pressure variation. However, there are more fundamental sources of randomness in all speech data that affect not only turbulent sounds such as sibilants, but also quasi-periodic sounds like vowels.

One source of this randomness is the recording equipment itself. The microphone, by its very nature as a physical sensor, is subject to small random changes in its behavior over time. Since the microphone mediates the physical sound wave and its record, these small random changes in the microphone's behavior engender random errors in the recorded data. Likewise, the recording device may introduce low-frequency background noise whose intensity varies randomly over time.

Moreover, it is known that speech is subject to intra-speaker variation. The waveforms of two tokens of the same word spoken by the same person, even in proximate succession, are assured to show unpredictable differences in their values, which may be due to differences in speaking rate, vocal effort, articulatory gestures, etc. that the speaker, much less the researcher, is unable to control from one production to the next; hence, this variation can be considered random.

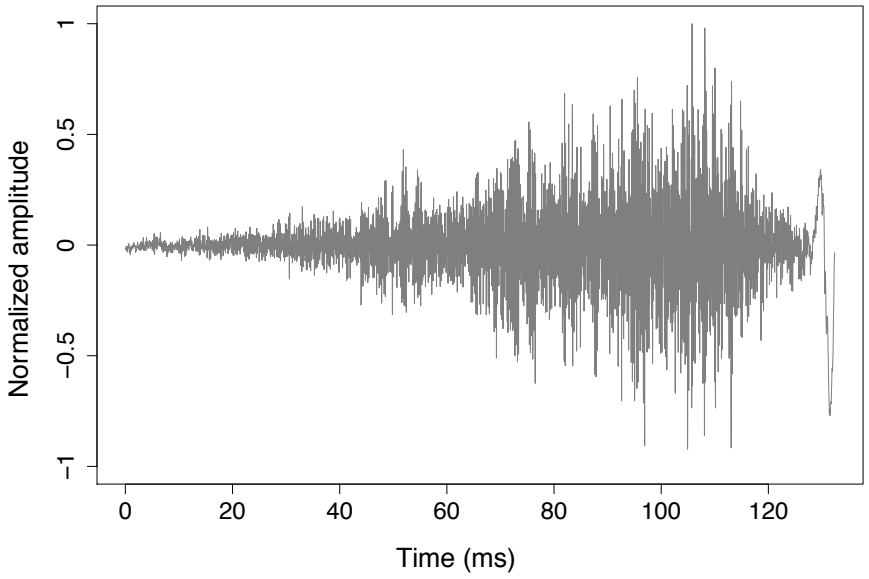

Figure 1: A realization of word-initial English /s/ excised from a token of 'sodas' as produced by an adult male speaker.

Due to the randomness intrinsic to speech data, each value of a waveform should be construed as a particular value taken by a random variable. For example, suppose that the researcher records a token of /s/ as the sequence of n numbers $x_1, x_2, \ldots, x_n$, where each value x_t is the t^{th} sampled value of the sound wave. Then, a natural model for the waveform of /s/ is a sequence of random variables $X_1, X_2, X_3, \ldots$, which just so happened to assume the values $x_1, x_2, \ldots, x_n$ when that particular token of /s/ was recorded. The decision to model the waveform of /s/ as a sequence of random variables correctly captures the fact that the values of the waveform of a token of /s/ are random in the sense discussed in the preceding two paragraphs, and motivates the introduction and definition of random processes.

Definition 1.1 (Random process)**.** A *random process* is a sequence of random variables, denoted by $\{X_t\}$, that are all defined on the same probability space, take values in the same measurable space, and are indexed by a variable t that ranges over (a subset of) the integers.

The measurable space in which each random variable takes its values is called the *state space* of the random process.

The linguistic objects suitable to be modeled by random processes are not limited to just the waveforms of phonetic segments. Indeed, the definition of a random process is a sequence of random variables, all of which are defined on a common probability space and share a common state space. The specifics of the probability space and state space are left open. So, all of the following could equally well be modeled by a random process: the waveform of a word, an f_0 track, a sequence of articulator positions, a text corpus. Each of

these examples reflects a change in the state space.

When a finite number of the variables in a random process $\{X_t\}$ assume values, the result is a sequence of numbers, referred to as a *realization* of the process and denoted by $\{x_t\}$; hence, the acoustic tokens of /s/ that form the data in the example are modeled as realizations of the random process that models /s/. So, when a random process's realizations are acoustic data, it should be clear that the random process models some waveform, not some sequence of states that describe constrictions in the vocal tract during the generation of that waveform, or some sequence of articulator postures that formed those vocal tract constrictions, or some sequence of motor unit activations that postured the articulators, or any other sequence of "articulatory states" at an even earlier point in the speech chain.

When a random process $\{X_t\}$ is used to model an acoustic waveform that has been sampled, the index t represents the (discrete) ordinal points in time at which the sound wave is sampled; hence, in the example of /s/, each random variable X_t models the t^{th} value of /s/'s waveform when sampled. If the sampling period T is known in seconds, then the "time" of each random variable in the random process can be given a physical meaning by associating each random variable X_t to the time tT seconds.

Once the researcher has collected a number of /s/ tokens, the spectral analysis can begin. For concreteness, suppose that the goal of the spectral analysis is to determine the peak frequency of the spectrum of /s/. From a procedural point of view, this analysis is straightforward: First, the spectrum of each /s/ token is computed; then, the peak frequency of each spectrum is determined; and finally, these values are used to estimate the peak frequency of /s/'s spectrum.

From a conceptual point of view, however, some elaboration is needed before this type of analysis can be considered meaningful. First, the notions of the "spectrum of /s/" and "the spectrum of a token of /s/" need to be clarified. Each of these ideas is resolved through the mathematical model of each linguistic object: A token of /s/ is modeled as a sequence of numbers, whose spectrum is known from the discrete Fourier transform (DFT). Therefore, the spectrum of a token of /s/ can be understood as the spectrum of the numeric sequence that models that /s/ token.

Likewise, the spectrum of /s/ refers to the spectrum of the random process that models /s/. But since the reach of traditional Fourier theory does not extend to random processes, this immediately exposes a hole in the logic of the procedure above. That is, the DFT is a map whose domain is a particular class of numeric sequence. This domain excludes all random processes; therefore, the DFT cannot be used to transform a random process into its spectrum. The change in mathematical object, whose spectrum is to be found, demands an extension of traditional Fourier theory. Without a theory of the spectral representation of random processes, a spectral analysis of /s/, or any other phonetic segment for that matter, is devoid of meaning.

In §2, the necessary extensions to traditional Fourier theory are reviewed. Specifically, it turns out that not all random processes have a spectral representation, so conditions on a

random process that guarantee the existence of its spectrum are presented. Furthermore, it is shown that each value of the spectrum of a random process $\{X_t\}$ depends on the infinite number of random variables in the process. However, $\{X_t\}$ is only ever observed as a realization of a finite number of variables; hence, each value of $\{X_t\}$'s spectrum can never be computed exactly. Instead, these values must be estimated from a finite realization.

In §3, two methods are presented for estimating the spectrum of a random process $\{X_t\}$ from a particular realization $x_1, x_2, \ldots, x_n$. Each method of estimation is evaluated analytically in order to explore how "close" to the true spectrum of $\{X_t\}$ an estimate computed from either method is expected to be. The form of these spectral estimators reveals the connection between the spectrum of $x_1, x_2, \ldots, x_n$ provided by the DFT and the spectrum of $\{X_t\}$. Specifically, both methods for estimating the spectrum of $\{X_t\}$ from $x_1, x_2, \ldots, x_n$ are based on the DFT of $x_1, x_2, \ldots, x_n$. This discussion, by extension, elucidates how the spectrum of a token of /s/ may be considered a representation of the spectrum of /s/.

Since the ultimate goal of the spectral analysis described above is the estimation of the peak frequency of /s/, rather than just its spectrum, the relationship between this or any other spectral property of /s/ and an estimate of it from a token of /s/ should be clarified as well. Mathematically, a spectral property of /s/ corresponds to a transformation of the spectrum of a random process $\{X_t\}$. For example, if the spectrum of $\{X_t\}$ is denoted by f_X, which ranges over a variable ω that denotes frequency, then the peak frequency of /s/ is given by the transformation

$$\mathsf{Peak}(X) = \arg\max_{\omega} f_X(\omega).$$

Similarly, a spectral property of a token of /s/ corresponds to a transformation of a spectral estimate computed from a realization $x_1, x_2, \ldots, x_n$. If this spectral estimate is denoted by S_x, which ranges over the discrete variable ω_j, then the peak frequency of the /s/ token is given by

$$\mathsf{Peak}(x) = \arg\max_{\omega_j} S_x(\omega_j).$$

While the discussion in §3 tells how S_x relates to f_X, it says nothing about how $\mathsf{Peak}(x)$ relates to $\mathsf{Peak}(X)$. The paper concludes with a discussion of the difficulties attendant with determining analytically how a spectral property of a random process $\{X_t\}$ relates to an estimate of that property computed from a realization $x_1, x_2, \ldots, x_n$. This difficulty of analysis implies that an analytic comparison of different methods for estimating a spectral property is for all practical purposes intractable. Instead, the researcher must justify which method of spectral estimation yields the "best" estimate of a given spectral property, by way of simulation rather than assuming that the relative merits of one spectral estimator over another transfer to estimates of spectral properties derived from that estimator.

2 Spectral representation of a random process

In this section, the theory of spectral representation for random processes is reviewed. The discussion is based on Shumway and Stoffer (2006), and the reader is referred there for a thorough general introduction to random processes and their spectral representation.

In the sequel, upper case letters $X, Y, \ldots$ are used to denote random variables; $\mathbb{E}(X)$ denotes the expected value of the random variable X; $\mathrm{Var}(X)$ denotes the variance of X; and $\mathrm{Cov}(X, Y)$ denotes the covariance between the random variables X and Y.[1] It is assumed that the reader is familiar with the meaning of all these terms.

In general, the methods from classical statistics cannot be applied to a random process because these methods assume that the random variables $\{X_t\}$ are independent and all follow the same distribution; however, a random process will not always obtain both of these properties. When used to model the waveform of a phonetic segment, the dependence structure of a random process and the change in its distributional properties over time are due to the nature of and physical constraints on speech production. First, speech production necessarily involves the movement of articulators, and as the posture of the articulators changes over time, the generated sound wave changes as well; hence, the distributional properties of a random process that models the wave form are expected to change with time as well. Second, the articulators move smoothly during the production of speech, and the posture that they can assume next depends on their current postural state. This dependence is projected forward in the speech chain, to the acoustic sound wave.

A complete description of the dependence structure of a random process $\{X_t\}$ would be had from knowing the joint cumulative distribution function of all finite subsets of random variables in $\{X_t\}$; however, such a complete description is usually unattainable. Instead, a much more limited description of $\{X_t\}$'s dependence structure is taken from its autocovariance function, which reports the covariance between each pair of variables in $\{X\}$. Below it is shown that the autocovariance function is intimately related to the spectral representation of a random process.

Definition 2.1 (Autocovariance function)**.** If $\{X_t\}$ is a random process, then the *autocovariance function* γ_X is defined by

$$\gamma_X(s, t) = \mathrm{Cov}(X_s, X_t). \tag{1}$$

In general, a process does not have a spectral representation; however, there is a very general subclass of random processes—the (weakly) stationary processes—which do admit such a frequency-domain representation.

[1] In general, it is possible that any of $\mathbb{E}(X)$, $\mathrm{Var}(X)$, or $\mathrm{Cov}(X, Y)$ may not converge to a value, making that value undefined; however, in the sequel it is assumed that all random variables have a finite expected value and variance and that all pairs of random variables have a finite covariance.

Definition 2.2 (Stationary process)**.** A random process $\{X_t\}$ is said to be *(weakly) stationary*[2] if it satisfies the following conditions:

1. $\mathbb{E}(X_t) = \mu$, for all t in the index set;
2. $\mathrm{Var}(X_t) < \infty$, for all t in the index set;
3. $\mathrm{Cov}(X_s, X_t) = \mathrm{Cov}(X_{s+h}, X_{t+h})$, for all s, t, $s+h$, and $t+h$ in the index set.

A process that does not satisfy the three conditions above is said to be *non-stationary*.

The third condition says that the covariance between any two random variables in a stationary process depends only on the amount of time that separates them, which allows its autocovariance function to be expressed in terms of a single variable denoting the separation between two random variables in the process: If $\{X_t\}$ is a stationary process with autocovariance function γ_X, then for all indices s and t with $h = s - t$, it follows that

$$\begin{aligned}\gamma_X(s,t) &= \gamma_X(s+h, t+h) \\ &= \gamma_X(s+h, s) \\ &= \gamma_X(h, 0),\end{aligned}$$

which does not depend on either time argument s or t. Hence, the autocovariance function of a stationary process can be expressed as a function of just the separation (or *lag*) h between two random variables,

$$\gamma_X(h) =_{def} \mathrm{Cov}(X_0, X_h). \tag{2}$$

If the $\{X_t\}$ models a waveform that is sampled with sampling period T seconds, then the lag h that separates two random variables in the process can be given the physical meaning of a separation of hT seconds.

The spectral representation of a stationary process can now be introduced. It can be proved that any stationary process can be expressed as a random linear combination of simple periodic functions oscillating at different frequencies (Shumway and Stoffer, 2006, Theorem C.2). Additionally, the autocovariance function of a stationary process also has a spectral representation, which is provided by the following theorem, stated without proof (Shumway and Stoffer, 2006, Property P4.1 & Theorem C.3).

Theorem 2.3 (Spectral Representation)**.** *If $\{X_t\}$ is a stationary process whose autocovariance function γ_X satisfies*

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty,$$

[2]By contrast, a process is said to be *strictly stationary* if the distributional properties of all finite subcollections of random variables in the process do not depend on time.

then there is a unique function f_X for which

$$\gamma_X(h) = \int_{-1/2}^{1/2} f_X(\omega) e^{2\pi i \omega h} \, \mathrm{d}\omega, \qquad h = 0, \pm 1, \pm 2, \ldots \tag{3}$$

The function f_X in (3) is called the spectral density *or* spectrum *of $\{X_t\}$ and is defined by*

$$f_X(\omega) = \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{-2\pi i \omega h}, \qquad \omega \in \mathbb{R}. \tag{4}$$

Readers who are familiar with traditional Fourier theory may notice that the spectral density f_X above is the Fourier transform of the (aperiodic, discrete) autocovariance function γ_X (Beerends et al., 2003, §18.5). This implies that f_X and γ_X uniquely determine each other, and that the spectral density f_X and the autocovariance function γ_X contain the same information since each value of γ_X can be recovered from f_X by integrating the right-hand side of (3). Therefore, we take the spectral density f_X of a stationary process $\{X_t\}$ as its foremost spectral representation and in the remainder of this section present some of the practical consequences of Theorem 2.3 for the spectral analysis of speech data.

2.1 Existence of a spectral representation of speech data

The first of these consequences concerns the speech data that a researcher is able to use to investigate spectral properties of a phonetic segment's waveform. Recall the example from the introduction, in which the waveform of /s/ is modeled by a random process $\{X_t\}$, and the data used in the study are modeled as realizations of $\{X_t\}$. In this setting, Theorem 2.3 implies that it is only meaningful to talk about the spectrum of the waveform of /s/ if that waveform is stationary. If /s/'s waveform is not stationary, then a stationary portion of /s/ must be isolated and used for the purposes of the spectral analysis.

Since the waveform of /s/ is only ever observed through a realization of it, this condition on the existence of a spectrum of (a portion of) /s/'s waveform, this raises the question of how to determine whether a particular token of /s/ is a realization of a stationary process. A rough but common method for checking this involves plotting the recorded token $x_1, x_2, \ldots, x_n$ as a function of time, and visually inspecting the mean and variance properties of its waveform. In particular, if the data is a realization of a stationary process, then it follows from definition 2.2(1) that the mean of $x_1, x_2, \ldots, x_n$ should be constant across time, and it follows from the following proposition that the variance should be constant as well.

Proposition 2.4 (Variance of a stationary process)**.** *If $\{X_t\}$ is a stationary process, then for every X_s and X_t in the process,* $\mathrm{Var}(X_s) = \mathrm{Var}(X_t)$.

Proof. Let X_s and X_t be random variables from a stationary process, and let $h = t - s$. Then, $\mathrm{Var}(X_s) = \mathrm{Cov}(X_s, X_s) = \mathrm{Cov}(X_{s+h}, X_{s+h}) = \mathrm{Cov}(X_t, X_t) = \mathrm{Var}(X_t)$.

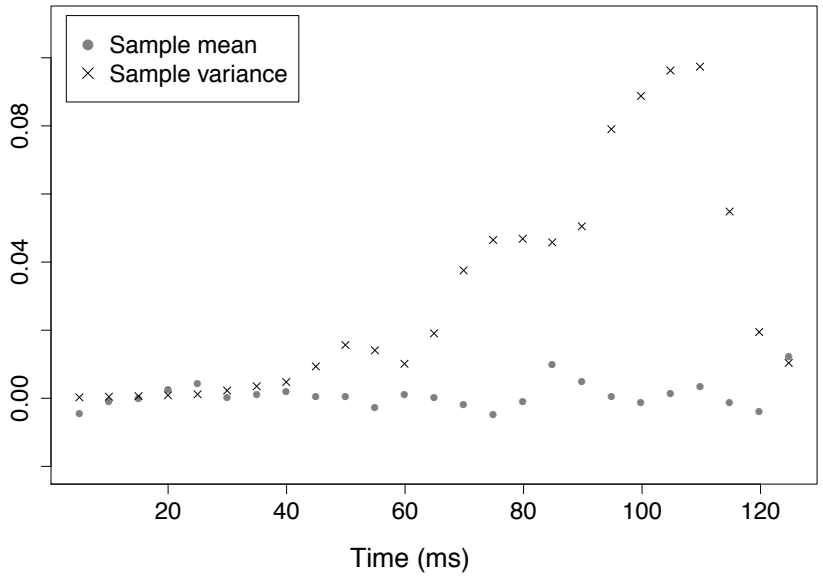

Figure 2: The sample mean (gray) and sample variance (black) of successive 10 ms data windows, with 5 ms overlap among adjacent windows, taken from the /s/ token shown in Figure 1. The value of each statistic is plotted against the time of the midpoint of the data window from which it was calculated.

The first and last equalities follow from the definition of covariance, and the second equality follows from the assumption that X_s and X_t come from a stationary process. □

Figure 2 shows the temporal progression of the sample mean and sample variance of successive 10 ms windows taken from the /s/ token shown in Figure 1. These statistics estimate the behavior of the evolution of the mean and variance of the random process $\{X_t\}$ that models /s/. From these plots, it is seen that the mean remains approximately constant, but the variance increases with time before decreasing sharply. So, this token of /s/ does not seem to be a realization of a stationary process, which, when considered in light of Theorem 2.3, suggests that it would be imprudent, much less meaningful, to use all the data from this token to estimate spectral properties of /s/.

In order to surmount this problem, the data can be used to hypothesize the location of a stationary subprocess of $\{X_t\}$, whose spectrum can be used as a proxy for that of the entire process. The data in Figure 2 suggest that the initial 40 ms interval of /s/ is stationary, as are the intervals between 40 and 60 ms and 70 and 90 ms. However, when automating a spectral analysis over a large data set, it is practically impossible to inspect each token individually in order to locate a stationary portion. Instead, it is common practice to take from each token a short interval placed in the same relative location, e.g. a 20 ms interval centered at the temporal midpoint of the waveform. It is taken on faith that the interval is of short enough duration that the statistical properties of the random process do not change

too drastically to violate the condition of stationarity. The fact that phoneticians typically restrict spectral analyses to "steady-state" portions of speech data suggests that random processes already occupy a very real, albeit unappreciated, role in phonetic analyses.

2.2 The domain of the spectral density function

In equation (4), the spectral density function f_X is defined over the entire real line; however, phoneticians are accustomed to visualizing the spectrum of speech data only on the interval of frequency values that ranges from 0 to the Nyquist frequency, $1/2T$ Hz, where T is the sampling period of the recorded data. Propositions 2.5 and 2.7 reconcile this discrepancy between theory and practice.

Proposition 2.5 (f_X is a periodic function)**.** *If $\{X_t\}$ is a random process that models an acoustic wave that is sampled with a sampling period T seconds, then its spectral density f_X is a periodic function with period $1/T$ Hz.*

Proof. From the discussion immediately following equation (2), each lag value h corresponds to hT units of time if $\{X_t\}$ models an acoustic wave that is sampled with sampling period T seconds. Therefore, equation (4) can be written as

$$f_X(\omega) = \sum_{h=-\infty}^{\infty} \gamma_X(hT) e^{-2\pi i \omega h T}, \tag{5}$$

where ω is expressed in Hz. Evaluating f_X at $\omega + 1/T$ then yields

$$\begin{aligned} f_X(\omega + 1/T) &= \sum_{h=-\infty}^{\infty} \gamma_X(hT) e^{-2\pi i(\omega + 1/T)hT} && (6)\\ &= \sum_{h=-\infty}^{\infty} \gamma_X(hT) e^{-2\pi i \omega h T} e^{-2\pi i h} && (7)\\ &= \sum_{h=-\infty}^{\infty} \gamma_X(hT) e^{-2\pi i \omega h T} && (8)\\ &= f_X(\omega). && (9) \end{aligned}$$

Equation (8) follows by virtue of the identity $e^{i\pi n} = 1$ for any integer n; in this case, $n = -2h$. □

The preceding proposition shows that the values of f_X are determined by the values that it takes on any interval whose size is $1/T$. Proposition 2.7 shows that the size of this interval can effectively be cut in half.

Lemma 2.6 (γ_X is an even function)**.** *If γ_X is the autocovariance function of a stationary process $\{X_t\}$ as defined by equation (2), then γ_X is an even function in the sense that $\gamma_X(-h) = \gamma_X(h)$ for all h.*

Proof. If γ_X is as described in the statement of the lemma, then

$$\begin{aligned}\gamma_X(h) &= \mathrm{Cov}(X_0, X_h) && (10)\\ &= \mathrm{Cov}(X_{-h}, X_0) && (11)\\ &= \mathrm{Cov}(X_0, X_{-h}) && (12)\\ &= \gamma_X(-h) && (13)\end{aligned}$$

Equation (11) follows from definition 2.2(3); equation (12), from the elementary fact that $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ for all random variables X and Y. □

Proposition 2.7 (f_X is an even function)**.** *If f_X is the spectral density function of a stationary process as defined by equation (4), then f_X is an even function.*

Proof. If f_X is the spectral density function of a stationary process, then

$$f_X(-\omega) = \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{2\pi i \omega h}. \tag{14}$$

Making the change of variable $h = -j$ yields

$$\begin{aligned}f_X(-\omega) &= \sum_{j=\infty}^{-\infty} \gamma_X(-j) e^{-2\pi i \omega j} && (15)\\ &= \sum_{j=\infty}^{-\infty} \gamma_X(j) e^{-2\pi i \omega j}. && (16)\end{aligned}$$

Equation (16) follows from Lemma 2.6. Comparison of equation (16) to equation (4) reveals that the former is just an alphabetic variant of the latter, where the summation is carried out in reverse. Therefore, it follows that $f_X(-\omega) = f_X(\omega)$, which proves the proposition. □

Taken together Propositions 2.5 and 2.7 show that if $\{X_t\}$ is a stationary process that models an acoustic wave sampled with sampling period T seconds, then its spectrum f_X is completely determined by the values that f_X takes on the frequency interval $[0, 1/2T]$ Hz. Since f_X is an even function, the values that it takes on the interval $[0, 1/2T]$ can be used to reconstruct its values on the interval $[-1/2T, 1/2T]$. The size of this reconstructed interval is $1/T$; hence, the values taken by f_X on this interval completely determine f_X on its entire domain since f_X is periodice with period $1/T$. Consequently, the spectrum of any given waveform need only be considered on the interval $[0, 1/2T]$, and in the following section all graphs of these spectra are shown only on this interval.

3 Spectral Estimation

In equation (4), each ordinate of the spectral density, $f_X(\omega)$, is expressed in terms of the autocovariance function γ_X; however, it is possible to express each ordinate in terms

of the random variables in $\{X_t\}$ by replacing γ_X in (4) with the righthand side of equation (2),

$$f_X(\omega) = \sum_{h=-\infty}^{\infty} \mathrm{Cov}(X_0, X_h) e^{-2\pi i \omega h}, \qquad \omega \in \mathbb{R}. \tag{17}$$

From this equation, it is immediately clear that the computation of each ordinate of f_X requires knowledge of the distributional properties of all the random variables in $\{X_t\}$; however, the random process is only ever observed as a finite realization $x_1, x_2, \ldots, x_n$. So, the value of each ordinate $f_X(\omega)$ cannot be computed exactly, but must instead be estimated.

A method for estimating the ordinates of f_X, which often takes the form of a function of a finite number of random variables $X_1, X_2, \ldots, X_n$ from a stationary process $\{X_t\}$, is referred to as a *spectral estimator*. The random variables $X_1, X_2, \ldots, X_n$ are called a *sample* of the process $\{X_t\}$. This section presents two spectral estimators that have been used in the spectral analysis of speech data: the windowed periodogram and the multitaper spectrum. Each of these spectral estimators finds its roots in the discrete Fourier transform (DFT), a spectral transform that is typically defined in terms of a finite numeric sequence (see Beerends et al. (2003, p. 360)). For the discussion of spectral estimators that follows, it is more convenient to define the DFT in terms of random variables rather fixed numbers.

Definition 3.1 (Discrete Fourier transform)**.** If $X_1, X_2, \ldots, X_n$ is a finite sequence of random variables from a stationary process $\{X_t\}$, then the *discrete Fourier transform* d_X of the sample is defined by

$$d_X(\omega_j) = \sum_{t=1}^{n} X_t e^{-2\pi i \omega_j t}, \tag{18}$$

where $\omega_j = j/n$ for $j = 0, \ldots, n-1$. The frequencies ω_j are referred to as the *Fourier frequencies*.

Other commonly encountered spectral transformations derived from the DFT are the *amplitude spectrum* $|d_X|$, defined by

$$|d_X|(\omega_j) = |d_X(\omega_j)|, \tag{19}$$

and the *power spectrum* $|d_X|^2$, defined by

$$|d_X|^2(\omega_j) = |d_X(\omega_j)|^2. \tag{20}$$

In equation (18), each ordinate of the DFT, $d_X(\omega_j)$ is defined as a sum of the random variables $X_1, X_2, \ldots, X_n$; hence, $d_X(\omega_j)$ is a univariate random variable since a sum of univariate random variables is itself a univariate random variable. Furthermore, each $d_X(\omega_j)$ estimates the value of $f_X(\omega_j)$, and as such is an example of a *point estimator*, i.e. an estimator of a single value. It follows that it is meaningful to investigate each ordinate's

distributional properties, such as its expected value and its variance. Knowlege of these properties for $d_X(\omega_j)$ enables a discussion of its *bias* and *mean square error (MSE)* as a point estimator. The latter is commonly used as a measure of the quality of a point estimator, so by extension the spectral estimators presented below can be compared via the MSE of their ordinates.

Definition 3.2 (Bias)**.** If $\hat{\theta}$ is a point estimator of a number θ, then the *bias* of $\hat{\theta}$, denoted $\beta\left(\hat{\theta}\right)$, is defined to be $\beta\left(\hat{\theta}\right) = \mathbb{E}\left(\hat{\theta}\right) - \theta$.

If $\beta\left(\hat{\theta}\right) = 0$, then $\hat{\theta}$ is said to be an *unbiased* estimator.

Definition 3.3 (Mean square error)**.** If $\hat{\theta}$ is a point estimator of a number θ, then the *mean square error* of $\hat{\theta}$, denoted $\mathsf{MSE}\left(\hat{\theta}\right)$, is defined to be $\mathsf{MSE}\left(\hat{\theta}\right) = \beta\left(\hat{\theta}\right) + \mathrm{Var}\left(\hat{\theta}\right)$.

The rest of this section is devoted to introducing and comparing the windowed periodogram and the multitaper spectrum. For each spectral estimator, its bias and variance are discussed only qualitatively; however, some comparison of the two estimators is still possible.

3.1 The windowed periodogram

The periodogram arises from scaling the power spectrum in (20) by the inverse of the number n of random variables available to the estimator.

Definition 3.4 (Periodogram)**.** If $X_1, X_2, \ldots, X_n$ are a sample from a stationary process $\{X_t\}$, then the *periodogram* I_X of the sample is defined by

$$I_X(\omega_j) = n^{-1}|d_X(\omega_j)|^2, \tag{21}$$

where j and ω_j are as they are in definition (3.1).

The periodogram is, in some sense, the most "direct" estimator of the spectral density f_X given a particular sample $X_1, X_2, \ldots, X_n$. To see why this is so, recall the definition of f_X from equation (4). One immediately apparent method for estimating $f_X(\omega)$ is to estimate the autocovariance function γ_X and then compute the DFT of the result. A common estimator of γ_X is the *sample autocovariance function* (Shumway and Stoffer, 2006, p. 30).

Definition 3.5 (Sample autocovariance function)**.** If $X_1, X_2, \ldots, X_n$ is a sample of a random process $\{X_t\}$, then the *sample autocovariance function* is defined by

$$\hat{\gamma}_X(h) = n^{-1}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X}), \tag{22}$$

where $\bar{X} = n^{-1}\sum_t X_t$ is called the *sample mean*.

It is possible to show that for Fourier frequencies other than $\omega_0 = 0$ the DFT of the sample autocovariance function is equal to the periodogram.

Proposition 3.6 (DFT of the sample autocovariance function)**.** *If $X_1, X_2, \ldots, X_n$ is a sample of a stationary process $\{X_t\}$, with sample autocovariance function $\hat{\gamma}_X$ and periodogram I_X, then for Fourier frequencies other than $\omega_0 = 0$,*

$$I_X(\omega_j) = \sum_{|h|<n} \hat{\gamma}_X(h) e^{-2\pi i \omega_j h}.$$

Proof. First note that for $\omega_j \neq 0$, the DFT can be written as[3]

$$d_X(\omega_j) = \sum_{t=1}^{n} (X_t - \bar{X}) e^{-2\pi i \omega_j t}, \tag{23}$$

Therefore, for Fourier frequencies other than ω_0 it follows that

$$I_X(\omega_j) = n^{-1}|d_X(\omega_j)|^2 = n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} (X_t - \bar{X})(X_s - \bar{X}) e^{-2\pi i \omega_j (t-s)} \tag{24}$$

$$= n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X})(X_t - \bar{X}) e^{-2\pi i \omega_j h} \tag{25}$$

$$= \sum_{|h|<n} \hat{\gamma}_X(h) e^{-2\pi i \omega_j h}. \tag{26}$$

Comparing equation (26) to definition 3.1, it is clear that the periodogram is equal to the DFT of the sample autocovariance function. □

This proposition establishes a nice parallel among the spectral representations of a stationary process and a sample of that process: The spectrum of each is the Fourier transform of its appropriate autocovariance function. In order to establish a more direct relationship between the spectral density and an estimator of it, the *windowed periodogram* is introduced.

Definition 3.7 (Windowed periodogram)**.** If $X_1, X_2, \ldots, X_n$ is a sample of a stationary process, and $w_1, w_2, \ldots, w_n$ is a sequence of numbers, then the *w-windowed periodogram* I_{wX} of the sample is defined by

$$I_{wX}(\omega_j) = n^{-1} \sum_{t=1}^{n} w_t X_t e^{-2\pi i \omega_j t}. \tag{27}$$

The sequence of numbers $w_1, w_2, \ldots, w_n$ is referred to as a *data window* or *data taper*.

[3]For any complex number $z \neq 1$, $\sum_{t=1}^{n} z^t = z(1 - z^n)/(1 - z)$. Let $\omega_j \neq 0$ and let $z = \exp(-2\pi i \omega_j)$. Then, $z \neq 1$ and $z^n = \exp(-2\pi i \omega_j n) = \exp(-2\pi i j n/n) = \exp(-2\pi i j) = 1$ since j is an integer; hence, $\sum_{t=1}^{n} z^n$ = z(1-1)/(1-z) = 0.

Then, to prove equation (23), expand the sum therein and cancel the $\bar{X} \sum_{t=1}^{n} e^{-2\pi i \omega_j t}$ term.

In Proposition 3.8 below, it is assumed that $X_1, X_2, \ldots, X_n$ is a sample from a zero-mean process $\{X_t\}$, meaning that $\mathbb{E}(X_t) = 0$, for each random variable X_t in the process. It is likely that a zero-mean process is a valid model for the acoustic waveform of speech because the sound waves generated during speech production travel as a chain of increases and decreases in air pressure, which are likely to cancel each other over time. Indeed, the data shown in Figure 2 suggest that the acoustic waveform of /s/ is well-modeled by a zero-mean process.

Proposition 3.8 (Expected value of periodogram ordinates)**.** *If $X_1, X_2, \ldots, X_n$ is a sample of a zero-mean stationary process $\{X_t\}$, and $w_1, w_2, \ldots, w_n$ is a data window, then the expected value of the w-windowed periodogram I_{wX} is*

$$\mathbb{E}[I_{wX}(\omega_j)] = \int_{-1/2}^{1/2} W_n(\omega_j - \omega) f_X(\omega)\ \mathrm{d}\omega, \tag{28}$$

where

$$W_n(\omega) = n^{-1} \left| \sum_{t=1}^{n} w_t e^{-2\pi i \omega t} \right|^2, \quad w \in \mathbb{R}. \tag{29}$$

W_n is called the kernel *of the data window $w_1, w_2, \ldots, w_n$.*

Proof. If the righthand side of (27) is expanded and one of the variables of summation changed to $h = t - s$, the result is

$$I_{wX}(\omega_j) = n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} X_t X_{t+|h|} e^{-2\pi i \omega_j h}.$$

Taking the expectation of both sides yields

$$\mathbb{E}[I_{wX}(\omega_j)] = n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} e^{-2\pi i \omega_j h} \mathbb{E}(X_t X_{t+|h|}) \tag{30}$$

$$= n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} e^{-2\pi i \omega_j h} \mathbb{E}[(X_t - \mathbb{E}(X_t))(X_{t+|h|} - \mathbb{E}(X_{t+|h|}))] \tag{31}$$

$$= n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} e^{-2\pi i \omega_j h} \gamma_X(h), \tag{32}$$

where the first equation follows from the linearity of the expected value operator; the second, from the fact that $\{X_t\}$ is a zero-mean process; and the third from equation (2).

Finally, substituting the righthand side of (3) for $\gamma_X(h)$ gives

$$\mathbb{E}[I_{wX}(\omega_j)] = \int_{-1/2}^{1/2} n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} e^{-2\pi i(\omega_j-\omega)h} f_X(\omega)\ \mathrm{d}\omega \tag{33}$$

$$= \int_{-1/2}^{1/2} n^{-1} \left| \sum_{t=1}^{n} w_t e^{-2\pi i(\omega_j-\omega)t} \right|^2 f_X(\omega)\ \mathrm{d}\omega \tag{34}$$

$$= \int_{-1/2}^{1/2} W_n(\omega_j - \omega) f_X(\omega)\ \mathrm{d}\omega, \tag{35}$$

where the last equation follows from (29). □

Proposition 3.8 shows that the w-windowed periodogram I_{wX} and the spectral density f_X are mediated by the kernel W_n of the particular data window used on the sample. More specifically, the integral on the righthand side of equation (35) says that the expected value of the estimator $I_{wX}(\omega_j)$ is found by taking the kernel W_n, "laying it on top" of f_X so that $W_n(0)$ coincides with $f_X(\omega_j)$, multiplying the values of each function that overlap, and then summing these products. This operation is called the *convolution* of W_n and f_X.

However, it is important to recognize that equation (35) does not describe how the windowed periodogram estimate of $f_X(\omega_j)$ is computed from a realization; that information is found in equation (27). Instead, equation (35) provides information about how the estimator $I_{wX}(\omega_j)$ would behave if a number of estimates of $f_X(\omega_j)$ were computed from different realizations and then averaged, which is a doorway to the bias of $I_{wX}(\omega_j)$.

3.1.1 Bias properties of the windowed periodogram

It should be clear from equation (35) that in order for the expected value of $I_{wX}(\omega_j)$ to be determined, it is necessary to know both the kernel W_n and the spectral density f_X; however, in applications involving speech data, it is rarely the case that anything is known about f_X since this would require knowledge of the distributional properties of the random process $\{X_t\}$ that models the speech data. It is possible that such distributional knowledge could become available from a complete theory of the aeroacoustics of speech production, but at the moment this theory is lacking. Consequently, the bias of each ordinate in the windowed periodogram, $\beta\left(I_{wX}(\omega_j)\right) = \mathbb{E}\left[I_{wX}(\omega_j)\right] - f_X(\omega_j)$, is unknown because both terms involved in its computation depend on f_X.

Since the direct computation of $\beta\left(I_{wX}(\omega_j)\right)$ is often impossible in practice, the bias properties of the windowed periodogram are explored through the kernel W_n, whose form depends on the particular window applied to the data before the spectral estimate is computed. This section discusses the kernel's of two data windows that should be familiar to phoneticians and other speech researchers: the rectangular window and the Hamming window.

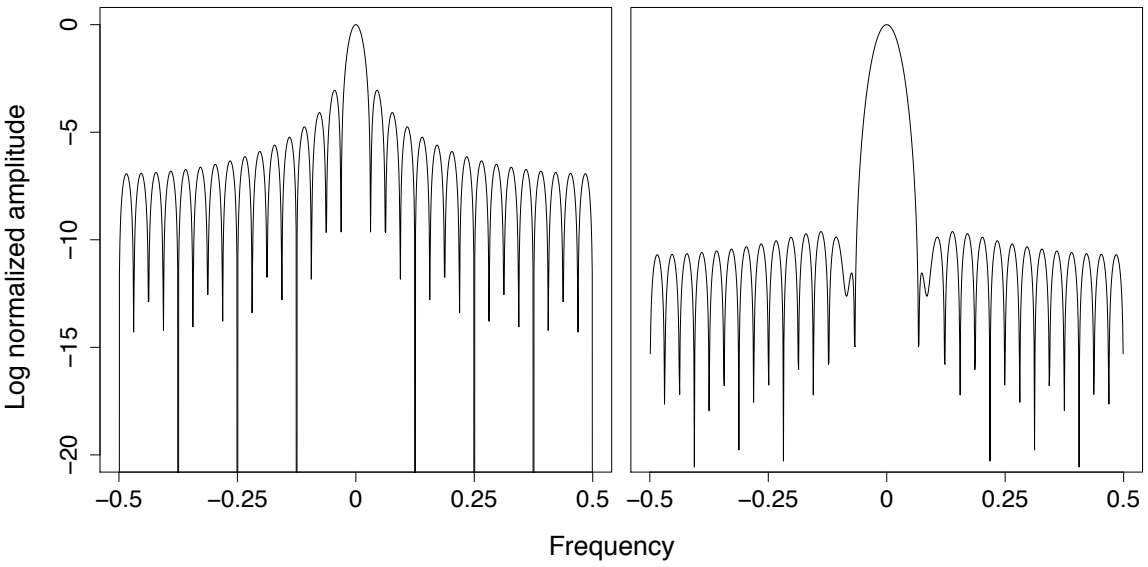

Figure 3: The kernel of the 32-point rectangular window (left panel) and the 32-point Hamming window (right panel). The values of each kernel were normalized by dividing by the maximum value of the kernel.

Definition 3.9 (Rectangular window)**.** The *(n-point) rectangular window* $r_1, r_2, \ldots, r_n$ is the sequence of n elements, each of which is equal to 1

Since $r_t X_t = X_t$ for $t = 1, 2, \ldots, n$, the rectangular window can be thought of as the "default" data window applied to the sample $X_1, X_2, \ldots, X_n$ when no other data window is used. From this it follows that the periodogram from Definition 3.4 is equal to the windowed periodogram from Definition 3.7 whose data window is the rectangular window; hence, the relationship between the windowed periodogram ordinate $I_{wX}(\omega_j)$ and the spectral density f_X established in Proposition 3.8 applies to the "unwindowed" periodogram as well, which implies that the bias properties of the periodram ordinates depend on the kernel of the rectangular window.

The kernel of the 32-point rectangular window is shown in the left panel of Figure 3. The shape of this kernel is characterised by a dominant peak, called the *main lobe*, centered at 0 with several other peaks, referred to collectively as the *side lobes*, on either side of it, whose respective heights decrease with their distance from the main lobe. While the number of side lobes in the kernel of a rectangular window depends on the length n of the window, the downward sloping pattern from the peak of the main lobe through the peaks of the side lobes is the same independent of n.

Consider how, according to equation (35), the shape of a kernel W_n affects the expected value, and by extension the bias, of $I_{wX}(\omega_j)$. The bias of $I_{wX}(\omega_j)$ is minimized when the

righthand side of this equation equals $f_X(\omega_j)$; however, when W_n and f_X are convolved, the value of f_X at each frequency $\omega \neq \omega_j$ is scaled by $W_n(\omega_j - \omega)$, and if $W_n(\omega_j - \omega) \neq 0$, then $W_n(\omega_j - \omega) f_X(\omega) \neq 0$ as well. Consequently, the degree to which $\mathbb{E}[I_{wX}(\omega_j)]$ is influenced by the value of the spectral density at a frequency $\omega \neq \omega_j$ is directly related to the magnitude of $W_n(\omega_j - \omega)$. Therefore, the height of the sidelobes of W_n gives some indication of the extent to which $\mathbb{E}[I_{wX}(\omega_j)]$ is corrupted by the values of f_X at frequencies different, and potentially far away, from $f_X(\omega_j)$, which in turn increases the magnitude of its bias. In sum, the height of the sidelobes of a kernel is a rough proxy measure of the bias of the spectral estimator related to that kernel—the greater the height of the kernel's sidelobes, the more biased the estimator.

The righthand panel of Figure 3 shows the kernel of the 32-point Hamming window.

Definition 3.10 (Hamming window)**.** The *n-point Hamming window* $h_1, h_2, \ldots, h_n$ is the sequence of numbers defined by

$$h_t = 0.5 \left(1 - \cos\left(\frac{2\pi(t-1)}{n-1}\right)\right), \quad t = 1, 2, \ldots, n. \tag{36}$$

The size of the sidelobes in the kernel of the Hamming window, relative to those in the rectangular window's kernel, suggests that the Hamming-windowed periodogram I_{hX} has better bias properties than the rectangular-window periodogram I_{rX}. Further support for this conclusion is provided by Figure 4, which shows a Hamming-window periodogram spectral estimate overlaid on a rectangular-window periodogram estimate, both of which were computed from the center 20 ms of the token of /s/ shown in Figure 1. Both estimates suggest that the most prominent peak of the spectral density occurs just below 5 kHz. Taking this together with Proposition 3.8 and both panels of Figure 3, it is expected that at high frequencies the values of the rectangular-windowed periodogram estimate would be higher than those of the hamming-windowed periodogram estimate since the sidelobes of the rectangular window's kernel are larger than those of the Hamming window's kernel.

While the differences in bias properties between the rectangular- and the Hamming-windowed periodogram are borne out by the example estimates in Figure 4, for the purposes of analyzing speech data, it is more important to focus on how or whether these differences affect the analysis rather than to focus on the purely theoretical concern as to whether they exist at all. For example, both windowed periodogram estimates share roughly the same shape; hence, if the analysis dictates that after the spectrum is estimated, its values are used to compute statistics that summarize its shape, e.g. the first four spectral moments, then it is not a foregone conclusion that the two spectral estimators will deliver different results, just by virtue of their having different bias properties.

3.1.2 Variance of the windowed periodogram

While the use of a data window such as the Hamming window can reduce the bias of each ordinate of the periodogram, the ordinates of the Hamming-windowed periodogram

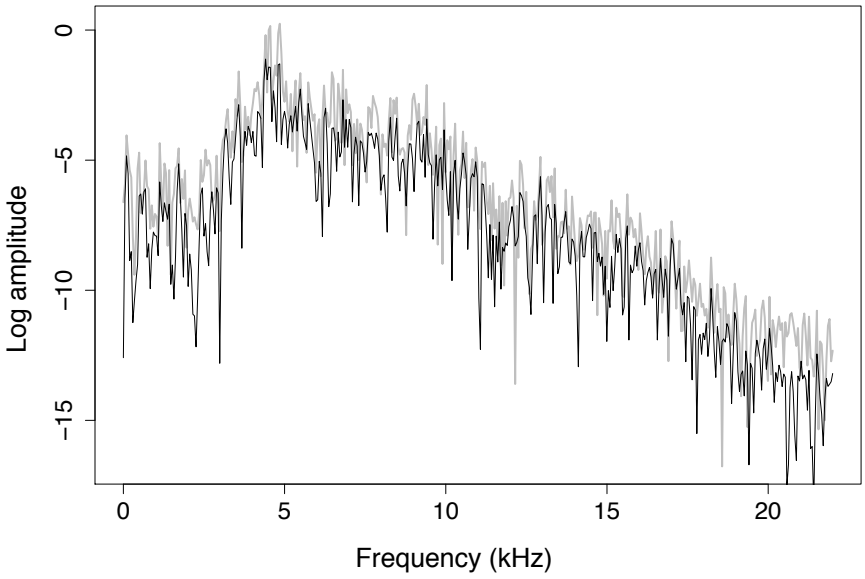

Figure 4: A comparison of a Hamming-window periodogram estimate (thin black line) and a rectangular-window periodogram estimate (thick gray line). Both estimates were computed from the center 20 ms of the token of /s/ shown in Figure 1.

I_{hX} are still prone to having a large MSE because of their large variance. The following theorem, based on Shumway and Stoffer (2006, p. 193, Property P4.2) and stated without proof, establishes the asymptotic distribution of each $I_{hX}(\omega_j)$, from which it is possible to investigate the variance of the estimator's ordinates.

Theorem 3.11 (Distribution of the windowed periodogram ordinates)**.** *If ω_j, $j = 0, 1, \ldots, n-1$, are distinct Fourier frequencies such that $f_X(\omega_j) \neq 0$, for all j, and if for each ω_j, $\{j_n\}$ is a sequence of integers such that $j_n/n \longrightarrow \omega_j$ as $n \longrightarrow \infty$, then as $n \longrightarrow \infty$,*

$$I_{hX}(j_n/n) \xrightarrow{d} \frac{f_X(\omega_j)}{2}\chi^2_2, \tag{37}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Hence, the variance of each ordinate of a Hamming-windowed periodogram is approximately

$$\begin{aligned} \operatorname{Var}\left[I_{hX}(\omega_j)\right] &\approx \operatorname{Var}\left[\frac{f_X(\omega_j)}{2}\chi^2_2\right] && (38)\\ &= \left(\frac{f_X(\omega_j)}{2}\right)^2 \operatorname{Var}[\chi^2_2] && (39)\\ &= f_X(\omega_j)^2. && (40) \end{aligned}$$

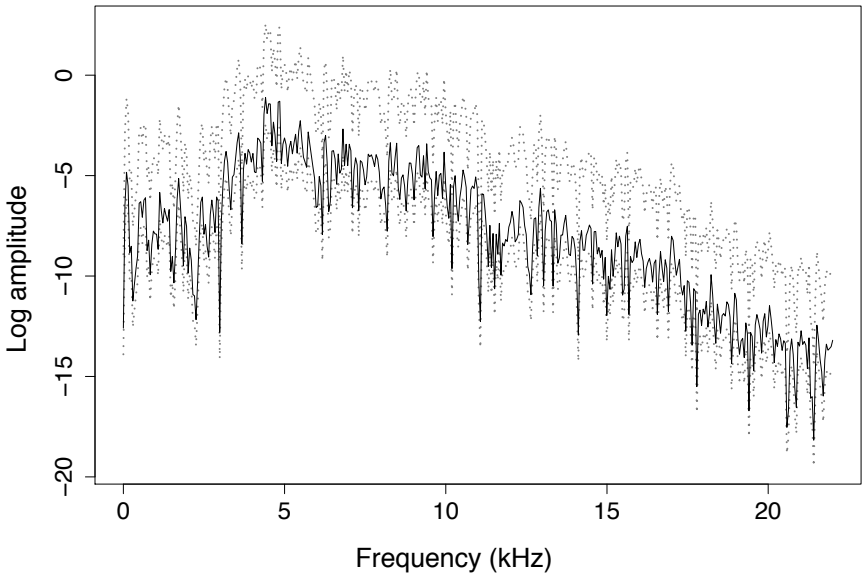

Figure 5: The Hamming-window periodogram estimate (black line) redrawn from Figure 4 plotted with the upper and lower bounds (gray dotted line) of a 95% confidence interval for each ordinate.

The asymptotic distribution of $I_{hX}(\omega_j)$ can also be used to approximate a confidence interval for $f_X(\omega_j)$ with confidence level $(1-\alpha)$. For a given α such that $0 < \alpha < 1$, under the asymptotic distribution of $I_{hX}(\omega_j)$ in (37), there is $(1-\alpha)$ probability that $I_{hX}(\omega_j)$ falls within the interval

$$\frac{f_X(\omega_j)}{2}\chi^2_2(\alpha/2) \leq I_{hX}(\omega_j) \leq \frac{f_X(\omega_j)}{2}\chi^2_2(1-\alpha/2),$$

where $\chi^2_2(\alpha)$, which is referred to as the *lower* α *probability tail*, is the number that satisfies $\mathbb{P}\left(\chi^2_2 < \chi^2_2(\alpha)\right) = \alpha$. Rearranging the terms in the above inequality yields a $100(1-\alpha)\%$ confidence interval for $f_X(\omega_j)$:

$$\frac{2I_{hX}(\omega_j)}{\chi^2_2(1-\alpha/2)} \leq f_X(\omega_j) \leq \frac{2I_{hX}(\omega_j)}{\chi^2_2(\alpha/2)}. \tag{41}$$

These confidence intervals can be visualized by plotting their upper and lower bounds against frequency. Figure 5 shows the Hamming-windowed periodogram estimate from Figure 4 along with the upper and lower bounds of a 95% confidence interval for each ordinate.

Furthermore, the form of the inequality in (41) suggests how the size of each confidence interval can be reduced. Specifically, the size of each confidence interval is directly related to the distance between the lower $(1-\alpha/2)$ and the lower $\alpha/2$ probability tails for the chi-squared distribution with two degrees of freedom. If it were possible to find a spectral estimator whose ordinates had an asymptotic distribution that depended on a distribution

δ whose variance was less than that of χ^2_2, then the distance between the lower $(1-\alpha/2)$ and the lower $\alpha/2$ probability tails of δ would be less than that for χ^2_2, and the size of the confidence interval for each ordinate of the spectral estimator would decrease as well. The desire for such a reduced-variance estimator motivates the introduction of the multitaper spectrum.

3.2 The multitaper spectrum

The multitaper spectrum was introduced by Thomson (1982), and the method of its calculation is simple enough: K copies of a sample $X_1, X_2, \ldots, X_n$ of a stationary process are weighted by K different data windows $\{w_{k,t}\}$. Then, for each windowed realization $\{w_{k,t}x_t\}$, its *eigenspectrum* S_k is found by computing its power spectrum. Finally, the K eigenspectra are averaged to produce the multitaper spectrum $M_X^{(K)}$.

It is also easy to get a sense for why this method would yield a spectral estimator, the variance of whose ordinates is less than that of the Hamming-windowed periodogram. If for a fixed Fourier frequency ω_j, the K ordinates $\{S_k(\omega_j)\}$ all have equal variance and are pairwise uncorrelated, then the ordinate of the multitaper spectrum at that frequency, $M_X^{(K)}(\omega_j)$, will have variance that is $1/K$ the size of the variance of $S_k(\omega_j)$. The aim is therefore to find data windows that will yield uncorrelated eigenspectra whose ordinates each have reasonable variance.

Data windows that satisfy these conditions are found in the family of discrete prolate spheroidal (DPS) sequences (Slepian and Pollak, 1961; Landau and Pollak, 1961, 1962; Slepian, 1964). These sequences were originally discovered as a solution to the spectral concentration problem, which asks whether it is possible to find a sequence of finite duration whose spectrum contains the maximal proportion of its energy in a fixed frequency band. To state the problem more concretely, the Fourier transform of a finite sequence is introduced (Beerends et al., 2003, § 18.5).

Definition 3.12 (Fourier transform of finite sequence)**.** If $x_1, x_2, \ldots, x_n$ is a finite sequence of real numbers, then its Fourier transform $\mathcal{X}$ is defined by

$$\mathcal{X}(\omega) = \sum_{k=1}^{n} x_k e^{-2\pi i \omega k}, \qquad -1/2 \leq \omega \leq 1/2. \tag{42}$$

If $x_1, x_2, \ldots, x_n$ is a sequence of length n, whose Fourier transform is $\mathcal{X}$, and W is a frequency such that $0 < W < 1/2$, then the *spectral concentration* of $\mathcal{X}$ in the band $[-W, W]$, denoted by $\lambda(n, W)$ is defined as

$$\lambda(n, W) = \frac{\int_{-W}^{W} |\mathcal{X}(\omega)|^2 \, \mathrm{d}\omega}{\int_{-\infty}^{\infty} |\mathcal{X}(\omega)|^2 . \, \mathrm{d}\omega} \tag{43}$$

The spectral concentration problem asks whether, given parameters n and W as above, it is possible to find the sequence that maximizes $\lambda(n, W)$. It turns out that the answer to this question is positive (Percival and Walden, 1993, Ch. 3 & 8). Moreover, it is possible to rank the sequences of length n according to their concentration $\lambda(n, W)$. This leads to the definition of a DPS sequence.

Definition 3.13 (DPS sequence)**.** Given fixed parameters n and W to the spectral concentration problem, the *DPS sequence of order k*, denoted by $\{v_t^{(k)}\}$ is the $(k+1)^{\text{th}}$ maximal concentration $\lambda(n, W)$.

So, the sequence that has the greatest energy concentration in the frequency band $[-W, W]$ is the DPS sequence of order 0; the sequence that has the second greatest energy concentration in $[-W, W]$ is the DPS sequence of order 1; and so on.

In order to generate DPS sequences that can be used as data windows for a spectral estimator is it necessary to set the frequency bandwidth parameter W. Conventionally, W is chosen so that the product nW is an integer that satisfies $nW \leq 4$. Futhermore, the choice of W places an upper bound on the number of eigenspectra K that are averaged to compute the multitaper spectrum. In particular, K should satisfy $K \leq 2nW$ (Percival and Walden, 1993, pp. 334-5).

As an illustration, DPS sequences were generated using the `multitaper` package for `R`, with the parameters $n = 883$ (the number of data points in a 20 ms waveform sampled at 44.1 kHz) and $W = 4/883$. For these parameters, the DPS sequences of orders $k = 0$ through $k = 5$ are shown in the top row of Figure 6. The corresponding eigenspectra for the center 20 ms of the /s/ from Figure 1 are shown in the bottom row of that same figure.

It can be shown that the ordinates of each eigenspectrum S_k all have the same asymptotic distribution as the ordinates of the Hamming-window periodogram (Percival and Walden, 1993, p. 343). That is, for all k such that $0 \leq k \leq K$ and j such that $0 \leq j \leq n-1$, the spectral ordinate estimator $S_k(\omega_j)$ converges in distribution to a scaled χ^2_2 random variable.

The DPS sequences are mutually *orthogonal*, in the sense that, for all orders j and k such that $j \neq k$,

$$\sum_{t=1}^{n} v_t^{(j)} \cdot v_t^{(k)} = 0,$$

which ensures that the eigenspectra used in the computation of the multitaper spectrum are pairwise uncorrelated (Percival and Walden, 1993).

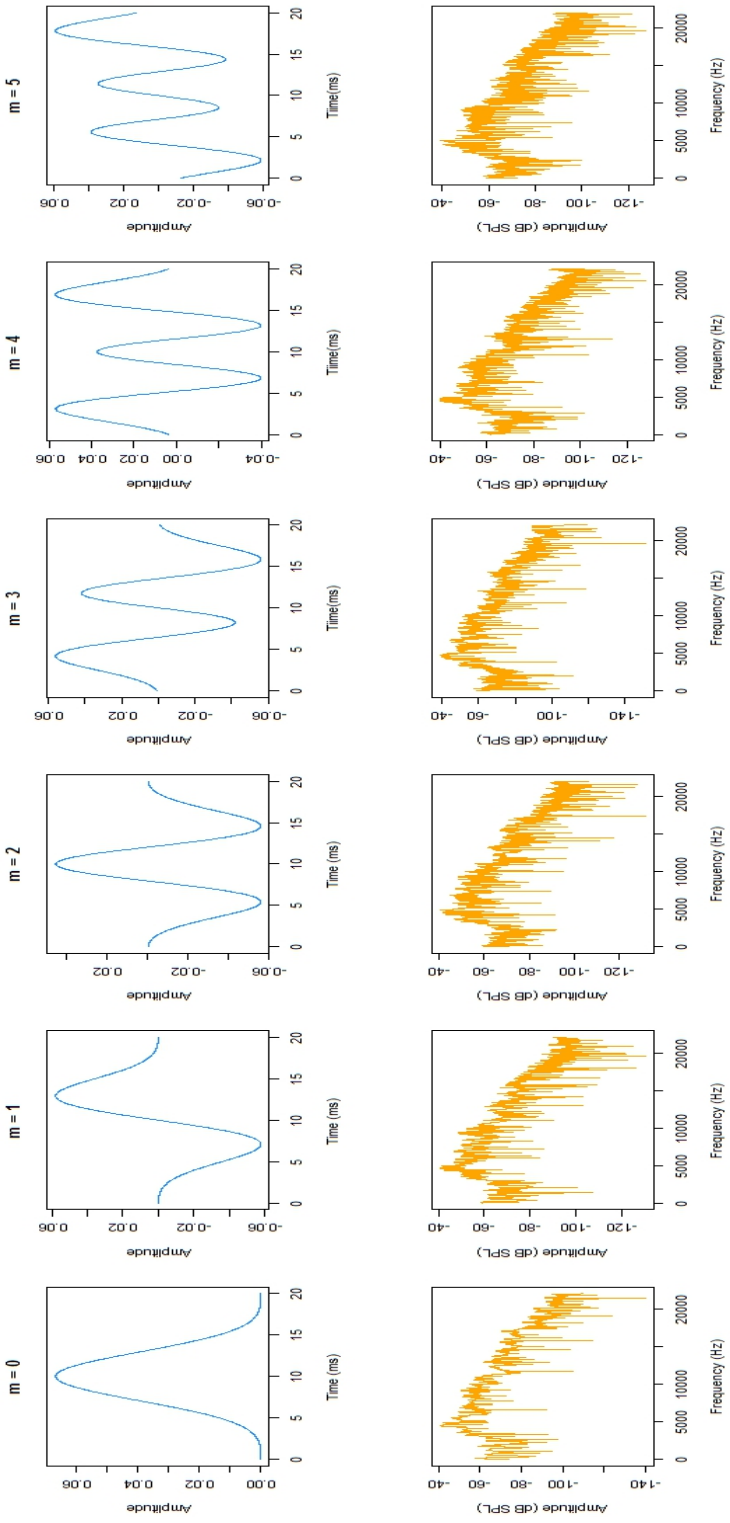

Figure 6: *Top row:* The DPS sequences of order $m = 0$ to $m = 5$, computed using the parameters $n = 883$ and $W = 4/883$. *Bottom row:* The eigenspectrum of the center 20 ms window of the /s/ from Figure 1, each computed using the DPS sequence above as a data window.

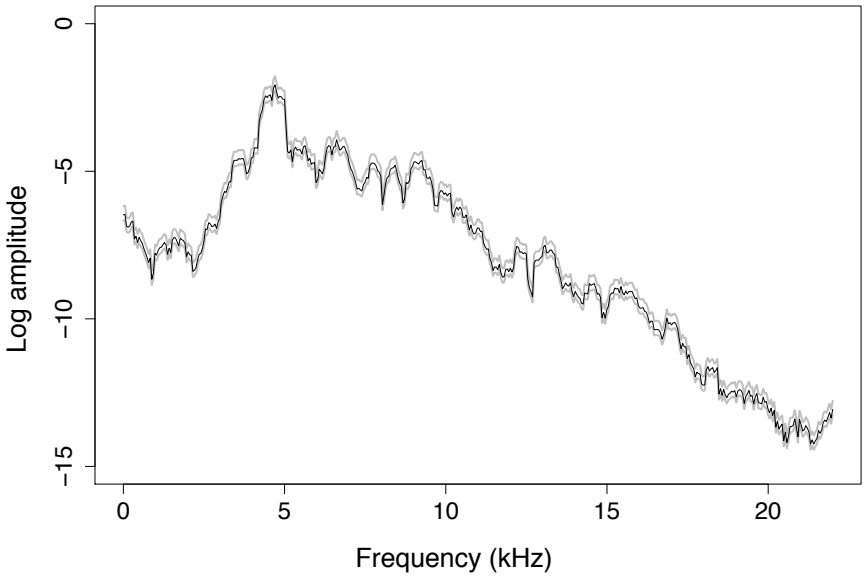

Figure 7: The multitaper spectrum (black line) of the center 20 ms of the /s/ from Figure 1, plotted with the upper and lower bounds (gray lines) of a 95% confidence interval for each ordinate.

The K mutually uncorrelated eigenspectra are averaged pointwise to compute the multitaper spectrum $M_X^{(K)}$; hence, the asymptotic distribution of each ordinate of $M_X^{(K)}$ is a scaled chi-squared with $2K$ degrees of freedom:

$$M_X^{(K)}(\omega_j) = \frac{1}{K}\sum_{k=0}^{K-1} S_k(\omega_j) \xrightarrow{d} \frac{f_X(\omega_j)}{2K}\chi^2_{2K}. \tag{44}$$

Using this distribution to approximate the variance of each ordinate $M_X^{(K)}(\omega_j)$ yields

$$\mathrm{Var}\left[M_X^{(K)}(\omega_j)\right] \;=\; \left(\frac{f_X(\omega_j)}{2K}\right)^2 \mathrm{Var}\left[\chi^2_{2K}\right] \;=\; \frac{f_x(\omega_j)^2}{K}. \tag{45}$$

Comparing (45) to (40), it is obvious that the ordinates of a multitaper spectrum have $1/K$ the variance of the ordinates of a Hamming-window periodogram.

The benefit of this reduced variance is revealed by the size of the confidence intervals for the ordinates of the multitaper spectrum. The confidence interval for each ordinate $M_X^{(K)}(\omega_j)$ is calculated analogously to (41),

$$\frac{2K\, M_X^{(K)}(\omega_j)}{\chi^2_{2K}(1-\alpha/2)} \le f_X(\omega_j) \le \frac{2K\, M_X^{(K)}(\omega_j)}{\chi^2_{2K}(\alpha/2)}. \tag{46}$$

Figure 7 shows the multitaper spectrum of the 20 ms of /s/ plotted with upper and lower bounds of a 95% confidence interval for each ordinate. The size of these confidence intervals is noticeably smaller than those in Figure 5: the mean size of the confidence intervals

for the ordinates of the Hamming-window periodogram estimate is 4.982, while the mean value of those of the multitaper spectrum estimate is 0.486. The correct interpretation of this difference in size is that when estimating an *ordinate* of the spectral density at a given frequency, $f_X(\omega_j)$, it is possible to circumscribe a smaller set of values within which $f_X(\omega_j)$ is likely to fall if the multitaper spectrum is used rather than the Hamming-window periodogram.

This brief introduction to spectral estimation has focused on two spectral estimators that have been used in speech applications: the Hamming-windowed periodogram and the multitaper spectrum. The comparison of these two estimators was carried out primarily in terms of the asymptotic variance of each estimator's ordinates, and it was shown that the multitaper spectrum has a variance that is a fraction of that of the Hamming-windowed periodogram.

4 Conclusion

In this paper, the spectral representation theory for random processes was reviewed, and two methods for estimating the spectrum of a random process were introduced and evaluated. The evaluation of the spectral estimators was carried out in theoretical terms, e.g. by comparing the variance of the asymptotic distribution of each estimator's ordinates. It was shown that the multitaper spectrum is a much "better" estimator than the Hamming-windowed periodogram in the sense that the variance of the former's ordinates is a fraction of that of the latter.

The reader is advised to bear in mind that this notion of "better" is purely theoretical, and in practice a spectral analysis of speech data usually does not end with the estimation of a spectrum, but with the estimation of *properties* of a spectrum, e.g. the peak frequency or one or more of the formants or spectral moments. Therefore, the comparison of the multitaper spectrum and the Hamming-windowed periodogram presented in §3 does not settle the question of which estimator is better-suited to a particular spectral analysis. In fact, it doesn't even address the question since doing so can only be done meaningfully once the details of the analysis are known.

References

Beerends, R. J.; H. G. ter Morsche; J. C. van den Berg; and E. M. van de Vrie. 2003. *Fourier and Laplace transforms*. Cambridge University Press, Cambridge, UK.

Landau, H. J., and H. O. Pollak. 1961. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—ii. *Bell System Technical Journal* 40.65–84.

Landau, H. J., and H. O. Pollak. 1962. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—iii: The dimension of the space of essentially time- and band-limited signals. *Bell System Technical Journal* 41.1295–1336.

Percival, Donald B., and Andrew T. Walden. 1993. *Spectral analysis for physical applications: Multitaper and conventional univariate techniques*. Cambridge University Press, Cambridge, UK.

Shumway, Robert H., and David S. Stoffer. 2006. *Time series analysis and its applications*. Springer Texts in Statistics. Springer, 2^{nd} edition.

Slepian, David. 1964. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—iv: Extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Technical Journal* 43.3009–3058.

Slepian, David, and H. O. Pollak. 1961. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—i. *Bell System Technical Journal* 40.43–64.

Thomson, David J. 1982. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70.1055–1096.

Appendices

The appendices below contain `R` code that can be used for the spectral analysis of speech data—in particular, the computation of Hamming-windowed periodogram and multitaper spectrum estimates.

A `Waveform.r`

```
# Author:       Patrick Reidy
# Affiliations: The Ohio State University
#               Department of Linguistics
#               www.ling.ohio-state.edu
#               Learning To Talk
#               www.learningtotalk.org
# Email:        reidy@ling.ohio-state.edu
# Mail:         24A Oxley Hall
#               1712 Neil Ave.
#               Columbus, OH 43210-1298
# License:      GPL-3


# The Waveform package depends on the Simon Urbanek's
# 'audio' package.
library('audio', quietly=TRUE)
```

```
########################
#  Utility functions  #
########################

`%@%` <- function(...) {
# %@% is a generic function for getting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%.default
  UseMethod('%@%')
}

`%@%.default` <- function(object, attribute) {
# %@%.default is the default method for getting the value of
# an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#                    slot of object.
# Returns:
#   The value of the R object in the attribute slot of
#   object.
  attr(object, attribute)
}

`%@%<-` <- function(...) {
# %@%<- is a generic function for setting the value of an
#   object's attribute(s).
# Methods available in the Waveform package:
#   %@%<-.default
  UseMethod('%@%<-')
}

`%@%<-.default` <- function(object, attribute, value) {
# %@%<-.default is the default method for setting the value
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#                    slot of object.
#       value: Any R object.
# Returns:
#   Nothing.  Instead, the value of object's attribute slot
#   is changed to value.
  `attr<-`(object, attribute, value)
```

```
}

.ConvertUnitNameToMultiplier <- function(unitName) {
# .ConvertUnitNameToMultiplier is a utility function for
# converting the time unit c('second', 'millisecond',
# 'microsecond', 'nanosecond') to proportions of one second.
# .ConverUnitNameToMultiplier implements the following map:
#        'second'   |-->   1
#   'millisecond'   |-->   1000
#   'microsecond'   |-->   1000000
#    'nanosecond'   |-->   1000000000
  if (unitName == 'second') {
    multiplier <- 1
  } else if (unitName == 'millisecond') {
    multiplier <- 1000
  } else if (unitName == 'microsecond') {
    multiplier <- 1000000
  } else if (unitName == 'nanosecond') {
    multiplier <- 1000000000
  }
  return(multiplier)
 }

.FindSampleAtTime <- function(waveform, timeOfSample,
                  timeUnit=(waveform %@% 'timeUnit')) {
# .FindSampleAtTime is a utility function for finding the
# index of the sample that occurs at a given time. Each
# sample point of the Waveform object is conceived of as
# being a half-open interval that is closed on the left and
# open on the right. The time value of each sample point
# is the value of the left boundary.
# Arguments:
#       waveform: A Waveform object.
#   timeOfSample: A numeric specifying the time of the
#                 sample whose index is to be found.
#       timeUnit: A character string specifying the unit of
#                 the timeValue argument.  Legal values:
#                 c('second', 'millisecond', 'microsecond',
#                 'nanosecond').
#                 Default is the unit of time for the time
#                 values of the Waveform object.
# Returns:
#   An integer specifying the index of the sample of the
#   Waveform object that occurs at the time specified by
#   the timeOfSample argument.
```

```
  # Create a vector of the sample times for the Waveform
  # object.
  sample.times <- .ComputeSampleTimes(waveform)

  # Convert the timeOfSample to the same unit as the sample
  # times.
  wave.time.unit <-
   .ConvertUnitNameToMultiplier(waveform %@% 'timeUnit')
  sample.time.unit <-
   .ConvertUnitNameToMultiplier(timeUnit)
  time.of.sample <-
   timeOfSample * (wave.time.unit / sample.time.unit)

  # Find the sample times that are prior or equal to the
  # time.of.sample.
  prior.sample.times <-
   which(sample.times <= time.of.sample)

  # The sample index is the last sample whose time is prior
  # or equal to the time.of.sample; hence, the index is
  # equal to the length of prior.sample.times
  sample.index <- length(prior.sample.times)

  # Return the index of the sample.
  return(sample.index)
}

.ComputeSampleTimes <- function(waveform) {
# .ComputeSampleTimes is a utility function for computing
# the time values of the sampled values of the waveform--
# i.e., the values that are not zeroes padded at the end of
# waveform in the case when waveform has been zero-padded.
# Arguments:
#   waveform: A Waveform object.
# Returns:
#   An integer specifying the number of sampled values in
#   waveform.

  # Find the start time of the waveform.
  start.time <- waveform %@% 'startTime'

  # Find the end time of the waveform.
  end.time <- waveform %@% 'endTime'
```

```
  # Find the number of samples in the waveform.
  sample.n <- waveform %@% 'N'

  # Create a vector of sample times from the start time, end
  # time and number of samples.
  sample.times <-
   seq(from=start.time, to=end.time, length.out=sample.n)

  # Return the vector of sample times.
  return(sample.times)
}

############################
#  Object initialization  #
############################

Waveform <- function(...) {
# Waveform is a generic function for creating a Waveform
# object.
# methods available in Waveform package:
#   Waveform.audioSample
#   Waveform.character
  UseMethod('Waveform')
}

Waveform.audioSample <-
function(audioSample, startTime=0, timeUnit='second') {
# Waveform.audioSample is a method for initializing a
# Waveform object from an audioSample object.  The
# audioSample class is defined in the 'audio' package.
# Arguments:
#   audioSample: An audioSample object.
#     startTime: A numeric that specifies the time of the
#                first sampled value of the audioSample
#                object.  Default is 0.
#      timeUnit: A character string specifying the unit of
#                measurement for startTime and endTime.
#                Legal values: c('second', 'millisecond',
#                   'microsecond', 'nanosecond').
#                Default is 'second'.
# Returns:
#   A Waveform object, which is a numeric vector whose
#   values represent the sampled values of the waveform,
#   augmented with the following attributes:
#        bitRate: The bitrate of the sampled waveform.
```

```
#     sampleRate: The sampling rate of the sampled waveform.
#   samplePeriod: The sampling period of the sampled
#                 waveform.
#              N: The number of samples in the waveform,
#                 excluding those values that are added to
#                 the waveform for zero-padding.
#      startTime: The time at which the first value of the
#                 waveform was sampled.
#        endTime: The time at which the last value of the
#                  waveform was sampled.
#       duration: The duration of the waveform.
#       timeUnit: The unit of measurement for startTime,
#                 endTime, and duration.

  # The audioSample object is a numeric vector that has a
  # 'bits' attribute and a 'rate' attribute for the bit rate and
  # sampling rate, respectively.
  waveform <- as.numeric(audioSample)

  # Set the 'bitRate' attribute of waveform to the value of
  # the 'bits' attribute of audioSample.
  waveform %@% 'bitRate' <- audioSample %@% 'bits'

  # Set the 'sampleRate' attribute of waveform to the value
  # of the 'rate' attribute of audioSample.
  waveform %@% 'sampleRate' <- audioSample %@% 'rate'

  # Set the 'samplePeriod' attribute of waveform.
  waveform %@% 'samplePeriod' <-
    1 / (waveform %@% 'sampleRate')

  # Set the 'N' (number of samples attribute of waveform.
  waveform %@% 'N' <- length(waveform)

  # Set the 'startTime' attribute of waveform from the
  # startTime argument.
  waveform %@% 'startTime' <- startTime

  # Set the 'endTime' attribute of waveform.
  time.lag.from.start <-
   ((waveform %@% 'N') - 1) / (waveform %@% 'sampleRate')
  waveform %@% 'endTime' <-
   time.lag.from.start + (waveform %@% 'startTime')

  # Set the 'duration' attribute of waveform.  The
```

```
  # (waveform %@% 'samplePeriod')  term is added in the
  # calculation below because each sampled point is a
  # treated as a semi-open interval (closed on the left,
  # open on the right) of duration equal to one sample
  # period.
  sampled.time.range <-
   (waveform %@% 'endTime') - (waveform %@% 'startTime')
  waveform %@% 'duration' <-
   sampled.time.range + (waveform %@% 'samplePeriod')

  # Set the 'timeUnit' attribute of waveform from the
  # timeUnit argument.
  waveform %@% 'timeUnit' <- timeUnit

  # Set the class of waveform.
  class(waveform) <- 'Waveform'

  # Return the Waveform object.
  return(waveform)
}

Waveform.character <-
function(waveFilepath, startTime=0, timeUnit='second') {
# Waveform.character is a method for initializing a Waveform
# object from the file path of a .wav file.
# Arguments:
#   waveFilepath: A character string specifying either the
#                 absolute or relative file path of a .wav
#                 file.
#      startTime: A numeric that specifies the time of the
#                 first sampled value of the waveform
#                 pointed to by waveFilepath.  Default is 0.
#       timeUnit: A character string specifying the unit of
#                 measurement for startTime and endTime.
#                 Legal values: c('second', 'millisecond',
#                 'microsecond', 'nanosecond').
#                 Default is 'second'.
# Returns:
#   A Waveform object, which is a numeric vector whose
#   values represent the sampled values of the waveform,
#   augmented with the following attributes:
#        bitRate: The bitrate of the sampled waveform.
#     sampleRate: The sampling rate of the sampled waveform.
#   samplePeriod: The sampling period of the sampled
#                 waveform.
```

```
#      startTime: The time at which the first value of the
#                 waveform was sampled.
#        endTime: The time at which the last value of the
#                 waveform was sampled.
#       duration: The duration of the waveform.
#       timeUnit: The unit of measurement for startTime,
#                 endTime, and duration.

  # Create an audioSample object by loading the wave file
  # pointed to by waveFilepath.
  audioSample <- load.wave(waveFilepath)

  # Dispatch the Waveform.audioSample method.
  Waveform(audioSample, startTime, timeUnit)
}

########################################
#  Methods to R's generic functions  #
########################################

plot.Waveform <- function(waveform, xAxisUnit='millisecond',
   type='l', col='orange',
   xlab=sprintf('Time (%s)', xAxisUnit), ylab='', ...) {
# plot.Waveform is a method for visualizing a Waveform
# object.
# Arguments:
#   waveform: A Waveform object.
#  xAxisUnit: A character string specifying the unit of the
#             time points plotted along the x-axis. Legal
#             values: c('second', 'millisecond',
#            'microsecond', 'nanosecond').
#             Default is 'millisecond'.
#       type: The type of line used to plot the values of
#             waveform. This value is passed to the
#             graphical parameter 'type'.
#        col: The color of the line used to plot the values
#             of waveform. This value is passed to the
#             graphical parameter 'col'.
#       xlab: The label on the x-axis.  This value is passed
#             to the graphical parameter 'xlab'. Default is
#             'Time (<xAxisUnit>)', where <xAxisUnit> is
#             replaced by the value of the xAxisUnit
#             argument.
#       ylab: The label on the y-axis.  This value is passed
#             to the graphical parameter 'ylab'.  Default is to
```

```
#             have no label.
#        ...: Other graphical parameters.
# Returns:
#   A plot of the Waveform object.

  # Make a vector of the time points at which the waveform's
  # samples occur.
  sample.times <- .ComputeSampleTimes(waveform)

  # Convert the unit of sample.times.
  wave.time.unit <-
     .ConvertUnitNameToMultiplier(waveform %@% 'timeUnit')
  x.axis.unit <- .ConvertUnitNameToMultiplier(xAxisUnit)
  sample.times <-
     sample.times * (x.axis.unit / wave.time.unit)

  # Grab just the sampled values of the waveform, excluding
  # the zero-padded values.
  wave.values <- as.numeric(waveform)
  sample.values <- wave.values[1:(waveform %@% 'N')]

  # Plot the Waveform object.
  plot(x=sample.times, y=sample.values,
       type=type, col=col, xlab=xlab, ylab=ylab, ...)
}

print.Waveform <- function(waveform) {
# print.Waveform is a method for reporting the attributes
# and visualizing a Waveform object.
# Arguments:
#   waveform: A Waveform object.
# Returns:
#   A report of the attributes of the waveform are printed
#   to the screen and a plot of the waveform is created.

  # Print the attributes of the Waveform object.
  message(sprintf(
     'Sampling rate:     %.2f', waveform %@% 'sampleRate'))
  message(sprintf(
     'Bit rate:          %d', waveform %@% 'bitRate'))
  message(sprintf(
     'Number of samples: %d', waveform %@% 'N'))
  message(sprintf(
     'Padded to:         %d', length(waveform)))
  message()
```

```
  message(sprintf(
     'Start time:        %f', waveform %@% 'startTime'))
  message(sprintf(
     'End time:          %f', waveform %@% 'endTime'))
  message(sprintf(
     'Duration:          %f', waveform %@% 'duration'))
  message(sprintf(
     'Time unit:         %s', waveform %@% 'timeUnit'))

  # Visualize the Waveform object.
  plot(waveform)
}

################################################
#  New generic functions and Waveform methods  #
################################################

FirstDifference <- function(...) {
  UseMethod('FirstDifference')
}

FirstDifference.Waveform <-
   function(waveform, coefficient=1) {
  # Make a delayed copy of the waveform that is scaled by
  # the coefficient.
  delayed.and.scaled <-
     c(0, waveform[1:(length(waveform)-1)]) * coefficient

  # Subtract the delayed and scaled copy from the waveform.
  preemphed.wave <- waveform - delayed.and.scaled

  # Return the pre-emphasized waveform.
  return(preemphed.wave)
}

TimeSlice <- function(...) {
# TimeSlice is a generic function for slicing a portion of a
# time series-like object according to time values, rather
# than indices.
# Methods available in Waveform package:
#   TimeSlice.Waveform
  UseMethod('TimeSlice')
}

TimeSlice.Waveform <- function(waveform, sliceFrom, sliceTo,
```

```
      centered=FALSE, duration,
      sliceUnit=(waveform %@% 'timeUnit')) {
# TimeSlice.Waveform is a method for slicing a portion of a
# Waveform object according to time values, rather than
# indices.
# Arguments:
#    waveform: A Waveform object.
#   sliceFrom: A numeric specifying the starting time of
#              the sliced portion of the Waveform object.
#     sliceTo: A numeric specifying the end time of the
#              sliced portion of the Waveform object.
#    centered: A boolean value.  If FALSE, then the values
#              of the sliceFrom and the sliceTo arguments
#              are used, and the duration argument is
#              ignored. If TRUE, then the duration argument
#              is used and the center portion of that
#              duration is sliced.
#    duration: A numeric specifying the duration of the
#              portion to be sliced if centered=TRUE.
#   sliceUnit: A character string specifying the unit of
#              time used to specify the sliceFrom and
#              sliceTo times. Legal values: c('second',
#              'millisecond', 'microsecond', 'nanosecond')
#              Default is the same time unit as the Waveform
#              object.
# Returns:
#   The portion of the Waveform object that falls between
#   the sliceFrom and sliceTo times, or a center portion of
#   the Waveform object if centered=TRUE.  If the Waveform
#   object had been zero-padded, then the zero-padding is
#   not appended to the sliced portion of the Waveform
#   object.

  # If the sliced portion is determined by sliceFrom and
  # sliceTo...
  if (! centered) {
    # Find the sample that occurs at the sliceFrom time.
    slice.from.index <- .FindSampleAtTime(waveform,
       timeOfSample=sliceFrom, timeUnit=sliceUnit)

    # Find the sample that occurs at the sliceTo time.
    slice.to.index <- .FindSampleAtTime(waveform,
       timeOfSample=sliceTo, timeUnit=sliceUnit)

    # Slice the waveform using the slice.from and slice.to
```

```
  # indices.
  sliced.wave <- as.numeric(waveform)
  sliced.wave <-
     sliced.wave[slice.from.index:slice.to.index]
} else {
# If the sliced portion is taken from the center of the
# waveform...
  # Compute the time of the midpoint of the waveform, in
  # waveform time units.
  wave.midpoint <- waveform %@% 'startTime' +
     ((waveform %@% 'duration') / 2)

  # Convert wave.midpoint from waveform time units to the
  # time units in which the slice duration is specified.
  wave.unit.factor <- .
     ConvertUnitNameToMultiplier(waveform %@% 'timeUnit')
  slice.unit.factor <-
     .ConvertUnitNameToMultiplier(sliceUnit)
  conversion.factor <-
     slice.unit.factor / wave.unit.factor
  wave.midpoint <- wave.midpoint * conversion.factor

  # Compute the time at the beginning of the sliced
  # portion.
  slice.from.time <- wave.midpoint - (duration / 2)

  # Find the sample that occurs at slice.from.time.
  slice.from.index <- .FindSampleAtTime(waveform,
     timeUnit=sliceUnit, timeOfSample=slice.from.time)

  # Compute the time at the end of the sliced portion.
  slice.to.time <- wave.midpoint + (duration / 2)

  # Find the sample that occurs at slice.to.time.
  slice.to.index <- .FindSampleAtTime(waveform,
     timeOfSample=slice.to.time, timeUnit=sliceUnit)

  # Slice the waveform using the slice.from and
  # slice.to indices.
  sliced.wave <- as.numeric(waveform)
  sliced.wave <-
     sliced.wave[slice.from.index:slice.to.index]
}

# Copy the attributes of waveform over to those of
```

```
  # sliced.waveform.
  attributes(sliced.wave) <- attributes(waveform)

  # Update the 'startTime', 'endTime', 'duration', and 'N'
  # attributes of sliced.wave.
  # First, create a vector of the sample times for the
  # unsliced Waveform object.
  sample.times <- .ComputeSampleTimes(waveform)
  # Second, set the 'startTime' attribute of sliced.wave to
  # the time of the first sliced sample.
  sliced.wave %@% 'startTime' <-
     sample.times[slice.from.index]
  # Third, set the 'endTime' attribute of sliced.wave to the
  # time of the last sliced sample.
  sliced.wave %@% 'endTime' <- sample.times[slice.to.index]
  # Fourth, calculate the duration of sliced.wave.
  sliced.range <- (sliced.wave %@% 'endTime') -
     (sliced.wave %@% 'startTime')
  sliced.wave %@% 'duration' <- sliced.range +
     (sliced.wave %@% 'samplePeriod')
  # Lastly, update the 'N' attribute of sliced.waveform.
  sliced.wave %@% 'N' <- length(sliced.wave)
  # Return the sliced waveform.
  return(sliced.wave)
}

Zeropad <- function(...) {
# Zeropad is a generic function for padding zeroes to the
# end of a time series-like object.
# Methods available in Waveform package:
#   Zeropad.Waveform
  UseMethod('Zeropad')
}

Zeropad.Waveform <-
   function(waveform, lengthOut=(waveform %@% 'sampleRate')) {
# Zeropad.Waveform is a method for padding zeroes to the
# end of a Waveform object.
# Arguments:
#    waveform: A Waveform object.
#   lengthOut: An integer specifying the length of the
#              Waveform object after if has been padded with
#              zeroes.  Default is to pad the waveform to
#              the length equal to its sampling rate.
# Returns:
```

```
#   A Waveform object that is identical to the original
#   Waveform object, but with zeroes added to the end of it.

  # Check that the lengthOut of the padded waveform is
  # greater than the number of sampled values in the
  # waveform.
  if ((waveform %@% 'N') < lengthOut) {
    # If so, pad the waveform.
    # First, make a copy of just the sampled values of the
    # waveform.
    wave.values <- as.numeric(waveform)
    sample.values <- wave.values[1:(waveform %@% 'N')]
    # Second, create a vector of 0's to pad to the end of
    # the sampled values.
    num.zeroes.to.pad <- lengthOut - (waveform %@% 'N')
    zeroes.to.pad <- rep(0, times=num.zeroes.to.pad)
    # Third, pad the zeroes to the sampled values.
    padded.waveform <- c(sample.values, zeroes.to.pad)
    # Lastly, copy the attributes of the Waveform object
    # over to the padded waveform.
    attributes(padded.waveform) <- attributes(waveform)
  } else {
    # If the number of sampled values is greater than the
    # length that the waveform should be padded to, then
    # it cannot be padded.
    padded.waveform <- waveform
    # Print an error message.
    message(
       'You must pad the waveform to a length that is')
    message(
       'number of sampled values in the waveform.')
    message()
    message(sprintf(
       'Number of sampled values: %d', waveform %@% 'N'))
  }

  # Return the padded waveform.
  return(padded.waveform)
}
```

B Tapers.r

```
# Author:       Patrick Reidy
# Affiliations: The Ohio State University
```

```
#               Department of Linguistics
#               www.ling.ohio-state.edu
#               Learning To Talk
#               www.learningtotalk.org
# Email:        reidy@ling.ohio-state.edu
# Mail:         24A Oxley Hall
#               1712 Neil Ave.
#               Columbus, OH 43210-1298
# License:      GPL-3

########################
#  Utility functions  #
########################

`%@%` <- function(...) {
# %@% is a generic function for getting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%.default
  UseMethod('%@%')
}

`%@%.default` <- function(object, attribute) {
# %@%.default is the default method for getting the value
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
# Returns:
#   The value of the R object in the attribute slot of
#   object.
  attr(object, attribute)
}

`%@%<-` <- function(...) {
# %@%<- is a generic function for setting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%<-.default
  UseMethod('%@%<-')
}

`%@%<-.default` <- function(object, attribute, value) {
# %@%<-.default is the default method for setting the value
```

```
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
#       value: Any R object.
# Returns:
#   Nothing. Instead, the value of object's attribute slot
#   is changed to value.
  `attr<-`(object, attribute, value)
}

############################
#  Hamming taper methods  #
############################

Hamming <- function(...) {
  UseMethod('Hamming')
}

Hamming.Waveform <- function(waveform) {
  # Get the number of samples in the Waveform object.
  num.of.samples <- waveform %@% 'N'

  # Generate the sequence of indices for the samples of the
  # Waveform object, that is a sequence of integers from 0
  # to (num.of.samples - 1).
  n.values <- seq(from=0, to=(num.of.samples -1 ))

  # Compute the values of the Hamming window from the
  # sequence of n values.
  hamming.values <- 0.54 -
     (0.46 * cos((2*pi*n.values) / (num.of.samples - 1)))

  # Pad the values of the Hamming window with the same
  # number of 0's that pad the Waveform object.
  zero.pad <-
     rep(0, times=(length(waveform) - num.of.samples))
  hamming.values <- c(hamming.values, zero.pad)

  # Multiply the Waveform object pointwise by the
  # zero-padded Hamming window.
  windowed.wave <- waveform * hamming.values

  # Set the attributes of the windowed waveform.
```

```
  attributes(windowed.wave) <- attributes(waveform)

  # Set an attribute to record how the waveform was
  # windowed.
  windowed.wave %@% 'taper' <- 'Hamming'

  # Return the windowed waveform.
  return(windowed.wave)
}
```

C Periodogram.r

```
# Author:       Patrick Reidy
# Affiliations: The Ohio State University
#               Department of Linguistics
#               www.ling.ohio-state.edu
#               Learning To Talk
#               www.learningtotalk.org
# Email:        reidy@ling.ohio-state.edu
# Mail:         24A Oxley Hall
#               1712 Neil Ave.
#               Columbus, OH 43210-1298
# License:      GPL-3

########################
#  Utility functions  #
########################

`%@%` <- function(...) {
# %@% is a generic function for getting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%.default
  UseMethod('%@%')
}

`%@%.default` <- function(object, attribute) {
# %@%.default is the default method for getting the value of
# an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
# Returns:
```

```
#   The value of the R object in the attribute slot of
#   object.
  attr(object, attribute)
}

`%@%<-` <- function(...) {
# %@%<- is a generic function for setting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%<-.default
  UseMethod('%@%<-')
}

`%@%<-.default` <- function(object, attribute, value) {
# %@%<-.default is the default method for setting the value of
# an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
#       value: Any R object.
# Returns:
#   Nothing. Instead, the value of object's attribute slot
#   is changed to value.
  `attr<-`(object, attribute, value)
}

############################
#  Object initialization  #
############################

Periodogram <- function(...) {
# Periodogram is a generic function for computing the
# ordinate values of the periodogram of a time series-like
# object.
  UseMethod('Periodogram')
}

Periodogram.Waveform <- function(waveform) {
# Periodogram.Waveform is a method for computing the
# ordinate values of the periodogram of a Waveform object.
# Arguments:
#   waveform: A Waveform object.
# Returns:
#   A Periodogram object, comprising the ordinate values of
```

```
#   the periodogram of the Waveform object.  If N is the
#   number of sampled values in the waveform x, then the
#   periodogram I of x is defined by
#     I(w_j) = (1/N) * |d(w_j)|^2,
#   where w_j = j/N is the j^th Fourier frequency (for j =
#   0, ..., N-1) and d(w_j) is the ordinate value of the
#   discrete Fourier transform at w_j, which is defined by
#     d(w_j) = \sum_{t=0}^{N-1} x_t * exp{-2 \pi i w_j t}.
#   A Periodogram object, furthermore, comprises the
#   following attributes:
#         nyquist: The Nyquist frequency of the Waveform
#                  object.
#               N: The number of sampled values in the
#                  Waveform object.
#        binWidth: The width of each frequency bin in the
#                  Periodogram object.
#    fourierFreqs: The hertz values of the Fourier
#                  frequencies.

  # Compute the ordinate values of the periodogram:
  # First, compute the ordinate values of the power spectrum.
  power.spectrum <- abs(fft(waveform))^2
  # Then, scale the power spectrum to get the periodogram.
  periodogram <- (1 / (waveform %@% 'N')) * power.spectrum

  # Keep only the ordinate values that lie on the upper half
  # of the unit circle.  That is, the ordinate values for
  # those frequencies that fall within [0, nyquist).
  nyquist.index <- floor(length(periodogram) / 2)
  periodogram <- periodogram[1:nyquist.index]

  # Set the 'nyquist' attribute of the periodogram, which is
  # equal to half the sampling rate of the Waveform object.
  periodogram %@% 'nyquist' <-
     (waveform %@% 'sampleRate') / 2

  # Set the 'N' attribute of the periodogram, which is equal
  # to the number of sampled values of the Waveform object.
  periodogram %@% 'N' <- waveform %@% 'N'

  # Set the 'binWidth' attribute of the periodogram, which
  # is equal to the sampling rate of the Waveform object,
  # divided by the number of values in the Waveform object
  # (including both sampled and zero-padded values).
  periodogram %@% 'binWidth' <-
```

```
      (waveform %@% 'sampleRate') / length(waveform)

  # Set the 'fourierFreqs' attribute of the periodogram.
  periodogram %@% 'fourierFreqs' <-
      seq(from=0, length.out=nyquist.index,
      by=(periodogram %@% 'binWidth'))

  # Set the class of the periodogram.
  class(periodogram) <-
      c('Periodogram', 'Spectrum', 'numeric')

  # Return the periodogram.
  return(periodogram)
}
```

D Multitaper.r

```
# Author:       Patrick Reidy
# Affiliations: The Ohio State University
#               Department of Linguistics
#               www.ling.ohio-state.edu
#               Learning To Talk
#               www.learningtotalk.org
# Email:        reidy@ling.ohio-state.edu
# Mail:         24A Oxley Hall
#               1712 Neil Ave.
#               Columbus, OH 43210-1298
# License:      GPL-3

# The Multitaper package depends on Karim Rahim's multitaper
# package.
library('multitaper', quietly=TRUE)

########################
#  Utility functions  #
########################

`%@%` <- function(...) {
# %@% is a generic function for getting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%.default
  UseMethod('%@%')
}
```

```
`%@%.default` <- function(object, attribute) {
# %@%.default is the default method for getting the value
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
# Returns:
#   The value of the R object in the attribute slot of
#    object.
  attr(object, attribute)
}

`%@%<-` <- function(...) {
# %@%<- is a generic function for setting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%<-.default
  UseMethod('%@%<-')
}

`%@%<-.default` <- function(object, attribute, value) {
# %@%<-.default is the default method for setting the value
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
#       value: Any R object.
# Returns:
#   Nothing.  Instead, the value of object's attribute slot
# is changed to value.
  `attr<-`(object, attribute, value)
}

.ColumnMultiply <- function(numVector, numMatrix) {
# .ColumnMultiply is a utility function for multiplying each
# column of a matrix by a vector.
# Arguments:
#   vect: A numeric vector.
#   matr: A numeric matrix.
# Returns:
#   A matrix that has the same dimensions as matr.  Each
#   column is equal to the corresponding column of matr
```

```
#   multiplied by vect.

  # Multiply each column of matr by vect.
  new.matrix <- apply(numMatrix, 2, `*`, numVector)

  # Return the new matrix.
  return(new.matrix)
}

.NormalizeSum <- function(numVector, normalizeTo=1) {
# .NormalizeSum is a utility function for normalizing the
# values of a numeric vector so that they sum to a
# predetermined number.
# Arguments:
#     numVector: A numeric vector.
#   normalizeTo: A numeric vector of length 1.
# Returns:
#   The numeric vector that results from scaling the
#   elements of numVector so that they sum to sumTo.

  # Determine the scale factor.
  scale.factor <- normalizeTo / sum(numVector)

  # Multiply by the scale factor.
  normalized.vector <- numVector * scale.factor

  # Return the normalized vector.
  return(normalized.vector)
}

############################
#  Object initialization  #
############################

Multitaper <- function(...) {
# Multitaper is a generic function for computing the
# multitaper spectrum of a time series-like object.
  UseMethod('Multitaper')
}

Multitaper.Waveform <- function(waveform, k=(2*nw), nw=4) {
# Multitaper.Waveform is a method for computing the
# multitaper spectrum of a Waveform object.
# Arguments:
#   waveform: A Waveform object.
```

```
#          k: An integer specifying the number of DPSS
#             tapers to use in the computation of the
#             multitaper spectrum.  Default value is
#             k = (2 * nw). Since k and nw are constrained
#             to satisfy k <= 2*nw, the default value is the
#             maximum number of tapers that should be used
#             given a fixed value of nw.
#         nw: An integer specifying the time-bandwidth
#             parameter used to generate the DPSS tapers.
# Returns:
#   The kˆth-order multitaper spectrum of the waveform,
#   computed using DPSS tapers generated using the
#   time-bandwidth parameter nw.  A Multitaper object,
#   furthermore, has the following attributes:
#         nyquist: The Nyquist frequency of the Waveform
#                  object.
#               N: The number of sampled values in the
#                  Waveform object.
#        binWidth: The width of each frequency bin in the
#                  Multitaper object.
#    fourierFreqs: The hertz values of the Fourier
#                  frequencies.
#               k: The number of DPSS tapers (equivalently,
#                  eigenspectra) used in the computation of
#                  the multitaper spectrum.
#              nw: The time-bandwidth parameter used to
#                  generate the DPSS tapers.

  # Generate the DPSS tapers using the dpss function from
  # the multitaper package.  dpss creates a named list whose
  # 'v' element is a matrix, each column of which is a DPSS
  # taper.
  dpss.taper.matrix <-
     dpss(n=(waveform %@% 'N'), k=k, nw=nw)$v

  # Zero-pad the DPSS tapers to the length of the waveform,
  # since the waveform is zero-padded by the difference
  # between length(waveform) and (waveform %@% 'N'):
  # First, determine how many zeroes were padded on the end
  # of the waveform.
  pad.length <- length(waveform) - (waveform %@% 'N')
  # Second, create a matrix of 0's that has pad.length rows
  # and k columns.
  zeropad.matrix <- matrix(data=0, nrow=pad.length, ncol=k)
  # Lastly, row bind the zeropad.matrix to the bottom of the
```

```
# dpss.taper.matrix.
dpss.taper.matrix <- r
   bind(dpss.taper.matrix, zeropad.matrix)

# Make k tapered copies of the sampled waveform by
# windowing it by each DPSS taper, using the
# .ColumnMultiply function.
tapered.wave.matrix <-
   .ColumnMultiply(waveform, dpss.taper.matrix)

# Compute the k eigenspectra of the waveform.  The k^th
# "eigenspectrum" of the waveform is the periodogram of
# the waveform after it has been windowed by the k^th
# DPSS taper.
eigenspectra <- abs(fft(tapered.wave.matrix))^2

# For each eigenspectrum, keep only the ordinate values
# for the frequencies in the [0, nyquist) range.
nyquist.index <- floor(length(waveform) / 2)
nyquist.eigenspectra <- eigenspectra[1:nyquist.index, ]

# Compute the k^th order multitaper spectrum by averaging
# the k eigenspectra, pointwise.
if (k == 1) {
  multitaper <- nyquist.eigenspectra
} else {
  multitaper <- rowMeans(nyquist.eigenspectra)
}

# Set the 'nyquist' attribute of the multitaper spectrum,
# which is equal to half the sampling rate of the
# Waveform object.
multitaper %@% 'nyquist' <-
   (waveform %@% 'sampleRate') / 2

# Set the 'N' attribute of the multitaper spectrum, which
# is equal to the number of sampled values of the Waveform
# object.
multitaper %@% 'N' <- waveform %@% 'N'

# Set the 'binWidth' attribute of the multitaper spectrum,
# which is equal to the sampling rate of the Waveform
# object, divided by the number of values in the Waveform
# object (including both sampled and zero-padded values).
multitaper %@% 'binWidth' <-
```

```
      (waveform %@% 'sampleRate') / length(wavefofrm)

  # Set the 'fourierFreqs' attribute of the multitaper
  # spectrum.
  multitaper %@% 'fourierFreqs' <-
     seq(from=0, length.out=nyquist.index,
     by=(multitaper %@% 'binWidth'))

  # Set the 'k' attribute of the multitaper spectrum.
  multitaper %@% 'k' <- k

  # Set the 'nw' attribute of the multitaper spectrum.
  multitaper %@% 'nw' <- nw

  # Set the class of the multitaper spectrum.
  class(multitaper) <- c('Multitaper', 'Spectrum', 'numeric')

  # Return the multitaper spectrum.
  return(multitaper)
}
```

AN ACOUSTIC ANALYSIS OF VOICING IN AMERICAN ENGLISH DENTAL FRICATIVES[1]

Bridget Smith
The Ohio State University

Abstract

In this study, an acoustic analysis of the dental fricatives as produced by American English speakers from the Buckeye Corpus (Pitt et al. 2006) reveals that the dental fricatives are subject to variation in voicing based on phonetic environment, much more than is usual for a pair of phonemes whose phonological distinction is based on voicing, confirmed by a comparison with the voicing of /f/ and /v/. The results of the study show that voicing (presence or absence of glottal pulses) for /θ/ and /ð/ is not predictable by phoneme in conversational speech, but it is more predictable based on voicing of surrounding sounds.

[1] This acoustic analysis was originally intended to be included in Smith 2009, "Dental fricatives and stops in Germanic." However, the editors and reviewers decided, and rightly so, that there were actually two papers there, so the acoustic analysis was removed and has been referred to as Smith 2007, "The seeds of sound change don't fall far from the tree," unpublished ms. While it has been my intention for some time to replicate the original study, introducing a number of new measurements and methods of analysis, while verifying the original measurements, this paper has sat on a back burner, still unpublished. While I still intend to perform such a reanalysis and submit it to a peer-reviewed journal, under whose scrutiny it will no doubt improve greatly, I submit this early version to the OSU Working Papers so that it may be more easily accessed by those who have expressed interest in it, for those who cannot wait for the next version to be published, and for those who have yet to discover it.

Bridget Smith. An acoustic analysis of voicing in American English dental fricatives. *OSUWPL*, Volume 60, Spring 2013, 117-128.

1. Introduction

The voiceless dental fricative (IPA /θ/) and the voiced dental fricative (IPA /ð/) are relatively understudied in comparison to other sounds. They are perceptually weak and easily confusable with /f/ and /v/. Because the voiced phoneme /ð/ appears word-initially in function words and word-finally in verbs (where the voiceless phoneme /θ/ does not appear), it is difficult to create psycholinguistic tasks in which the salience of the contrast can be measured. And because /θ/ and /ð/ may vary along multiple acoustic dimensions at once, forced choice tasks along a single dimension may not be particularly informative. Categorical replacement of /θ/ and /ð/ by /f/ and /v/ or /t/ and /d/ in certain dialects and sociolects has been studied, such as in AAVE (Wolfram 1970, 1974, among others) or London Cockney English (Wells 1982; Hughes, Trudgill & Watt 2005, among others). Polka, Colantonio, & Sundara (2001) found that English-speaking infants were less able to distinguish between /d/ and /ð/ than between /b/ and /v/, which suggests that the variation in production of the dental fricative which infants are exposed to overlaps to some extent with the alveolar (or dental) stop, so that they are unable to interpret a phonemic pattern until they are much older. A number of studies have looked at acoustic measurements to distinguish place of articulation among various groups of fricatives, and have either avoided /θ/ and /ð/, or were least successful in distinguishing /θ/ from /f/.

Stevens et al. (1992) show that the presence or absence of phonation is the acoustic parameter that best distinguishes between voiced and voiceless fricatives, and they are by no means alone in this judgment. See, e.g., Pirello, Blumstein, & Kurowski (1997). Denes (1955), however, demonstrated earlier that duration of a word-final fricative, and comparatively, the duration of the preceding vowel, could be manipulated to give the impression of voicing for longer vowels and shorter fricatives, and of voicelessness for shorter vowels and longer fricatives. Raphael (1972) confirmed these findings, and noted that "when the voicing characteristic is cued by vowel duration, perception is continuous rather than categorical" (1296). Pirello et al. (1997) say also that the production aspect of voicing is itself continuous, in that "the feature voicing in fricatives is manifested in a continuous way and as such cannot be characterized in terms of a binary distinction relating to the presence or absence of glottal excitation" (3754). Due to the complex nature of producing a voiced fricative, either voicing or frication may be lost during production: too little supraglottal pressure and frication fails, but if the supraglottal pressure is not sufficiently less than the subglottal pressure, voicing will fail. In addition, coarticulation frequently results in gestural overlap, with voicing (or the lack thereof) spreading from neighboring segments. It is for these reasons that a binary presence or absence of glottal pulses is insufficient to discriminate between voiced and voiceless fricatives. The amount of voicing overlap and the issue of duration is greatest for medial and final fricatives, but Pirello et al. (1997) achieved 93% accuracy classifying word-initial /s/ and /z/, and /f/ and /v/ from read speech. Using Stevens et al.'s (1992) rubric, in which the amplitude of the first harmonic of the fricative was compared to that of the following vowel, the fricatives were categorized as voiced and voiceless. The voiceless label was assigned to tokens with 10 dB or greater difference in amplitude. They argued that presence of glottal excitation present in at least 30 ms of either the beginning or the

end of the fricative was enough to correctly distinguish voiced from voiceless fricatives; although, they did not examine /θ/ and /ð/.

These previous studies, and others, have relied upon lab-produced read speech. While this allows researchers to exert some measure of control over variation, and creates tokens that can be easily compared across speakers, it does not give us a real picture of what these fricatives look like in conversational speech, and does not give an accurate picture of how these sounds might be discriminated by language users. Accumulated anecdotal observations provided the questions for this study: How strong is the voicing distinction between /ð/ and /θ/ in conversational speech? And is this distinction based on phonation, or do duration, intensity, or even manner of articulation play a greater role in conversational speech? When does voicing occur? What phonetic factors may be related to voicing? Does the phonological description of 'phoneme' match up with the phonetic realizations of /ð/ and /θ/? A parallel examination of /f/ and /v/ was conducted to find out what measures might be significant in distinguishing the voiced from the voiceless segments, and to provide a control group for comparison.

2. History

Both the voiced and voiceless dental fricatives are represented orthographically in English by the digraph <th>. /θ/ usually occurs word-initially or word-finally, and can occur medially in loanwords and certain compounds. /ð/ occurs word-initially only in function words such as articles and demonstratives, rarely in word-final position in certain derivational words such as *bathe* or *teethe*, and occurs in medial position in a greater number of words. There are a few minimal pairs, such as *thigh* and *thy, either* and *ether,* and *teeth* and *teethe*, and some near-minimal pairs such as *breath* and *breathe.* Despite the existence of at least one minimal pair in all positions, the contrast between these phonemes carries little, if any, functional load. The minimal pairs that exist cannot be used in the same position in a sentence, belonging generally to different classes of words.

This distribution is easily explained through the historical development of these sounds. In Old English, the *thorn* <þ> and *edh* <ð> characters interchangeably represented both the voiced and voiceless variant, which were at that time in complementary distribution. It is generally assumed that thorn or edh in initial and final position was voiceless, while between voiced sounds, it was voiced. /f/ and /v/ (and /s/ and /z/) had a similar distribution. Early on in Middle English, /f/ and /v/ (and /s/ and /z/) became phonemic, due to a confluence of factors, not least of which was the introduction of large numbers of loanwords containing these sounds in contrastive positions. Late during the Middle English period, it was noticed that there were two <th> variants, presumably a voiced and a voiceless phoneme (Bullokar 1580). Function words such as *the, that, this, then*, etc., are assumed to have begun with a voiceless dental fricative in Old English, but their Modern English counterparts have become voiced. Because they are often unstressed and not discrete from adjacent words, they are more likely to assimilate to surrounding voiced sounds. Another possible contributing factor is that the high frequency of these function words may have allowed a large amount of variation,

which became generalized as a voicing contrast. Note that the phonologization of these sounds occurred after the paradigm leveling that reduced the number of different forms of these function words. For example, the definite article *the* was inflected for case, gender, and number in Old English, yielding approximately 12 distinct forms of this word. The increased frequency of single forms of certain types of words may have created the situation that allowed reduction and variation of these high frequency words that now carry much less grammatical information. Word-final /ð/ also appeared in late Middle or Early Modern English, with the loss of verb endings stranding the medially-voiced fricative at the end of the verb. The greatest number of instances of /θ/ that occur outside of the original conditioning environment are in more recent loanwords such as *author, arithmetic,* and *arthritis*, and in forms that have undergone some kind of analogy or reanalysis, such as *Arthur* or *anthem*. /v/ and /f/ have followed a similar trajectory, with fossilization of the original conditioning environment in many words, but with a much greater number of minimal pairs from borrowing, though the contrast is arguably less than that of /s/ and /z/. Therefore, it stands to reason that /v/ and /f/ are the closest point of comparison to /ð/ and /θ/.

3. Methods

Eight talkers (four men, four women) were selected from the Buckeye Corpus (Pitt et al. 2006), which is a body of 40 sociolinguistic interviews with Ohio residents. After subsequent analysis, one of the male talkers' data were excluded from this analysis because of speech differences possibly resulting from a head injury. From approximately 15 minutes of conversational speech per talker, around 400 /ð/ and /θ/ and 300 /v/ and /f/ tokens were measured altogether. Between 15 and 25 /θ/ tokens were measured from each speaker, 25-45 /ð/ tokens, 11-30 /f/ tokens, and 10-35 /v/ tokens. Additional tokens were marked as assimilated or deleted, but not included in the measurements.

In Praat (Boersma & Weenink 2007), an acoustic analysis software, intervals were measured for the duration of frication of /θ/, /ð/, /f/, and /v/. Average intensity of these periods was also measured. Because of the variability of conversational speech, and the greater degree of coarticulation, no single uniform cue existed for determining the beginning or end of the fricative. Obvious frication was used as a marker, where it was present, generally in voiceless tokens. In voiced fricatives, either frication or reduced amplitude of formants as compared to surrounding vowels was a reliable marker. In some cases, /ð/ could be perceived in the speech signal, but not identified by characteristics in the waveform or spectrogram. In these and other cases in which the sound was completely unable to be differentiated from surrounding sounds, it was marked as assimilated or deleted. The beginning and ending of periods of voicing within the fricative were also measured. Intervals were marked as voiced if periodicity in the waveform and/or presence of a voice bar in the spectrogram indicated regular glottal pulses. To minimize confusion with the phonological feature [+voice], these are described as having *voice bar*. A measure of voicing was then created by taking the percentage of the duration of the fricative in which a voice bar could be found. The immediately preceding and following segments adjacent to each fricative, and whether or not they were voiced, were marked and noted. Neighboring segments were only classified

as voiceless if there was no periodicity in the waveform for more than 50% of the segment. A neighboring pause also counted as a voiceless segment because the vocal folds were not vibrating.

A subset of four talkers was selected for an additional analysis examining manner of articulation of the dental fricatives as well. The author examined the waveform and spectrogram, and listened to each token to arrive at the manner judgments. A description of characteristics used to identify each manner is given in the appendix. A second trained transcriber also gave manner judgments for each segment. For the fewer than 5% of the judgments that did not match between the two transcribers, these were then judged in tandem and an appropriate label agreed on. The majority of cases were approximants or flaps that could not be decided on, and without any concrete way of separating these two, the categories were combined into one.

4. Results

The results of the manner judgments are listed in Table 1. /ð/ was realized as a voiced fricative only 20.3% of the time. 23.4% of the time it was realized as a nasal, 18.4% as a stop, and 15.4% as an approximant or flap. /θ/, on the other hand was realized as a voiceless fricative 55.1% of the time, but was also realized as a stop 15.9% of the time, and as a voiced fricative 10.1% of the time. Because of the lack of contrast in manner at the interdental/dental place of articulation, /ð/ and /θ/ are free to vary in this way, though there is much more variation than one might expect.

Table 1: Manner judgments of dental fricatives

Manner of articulation	**edh /ð/**	**theta /θ/**
affricate	0.013	0.029
approximant/flap	0.152	0.043
deletion/vowel	0.095	0.015
voiced fricative	0.203	0.101
voiceless fricative	0.088	0.551
fricative+approximant	0.006	0.0
fricative+stop	0.019	0.058
lateral	0.006	0.0
nasal	0.234	0.015
stop	0.184	0.159
stop+lateral	0.0	0.029

Tokens which were completely assimilated or deleted were not included in the acoustic measurements, because they could not be segmented away from the following sounds. An additional 2 /v/ tokens and 5 /θ/ tokens were removed as outliers, having durations greater than 3 standard deviations from the mean, and were greatly lengthened while the interviewee was thinking of what he/she wanted to say next.

Measurements of absolute duration, as in Figure 1, reveal a great amount of overlap, which is unsurprising for conversational speech. What is notable, however, is that /ð/ and /θ/ pattern very similarly, while /f/ and /v/ show a bimodal distribution, with /f/ having a shorter duration (median 11 ms) than /v/ (median 35 ms), which is entirely the opposite pattern from we would expect. /ð/ and /θ/, while having a similar distribution to each other, also have the majority of tokens clustered around a very short duration, with half the tokens of both being less than 15 ms. These tokens are initial, medial, and final, and in a variety of mono- and polysyllabic words with varying stress. Many neighboring segments were not vowels, and many vowels were reduced to the point of deletion. Calculating the fricative duration relative to surrounding vowel duration did not make sense in this case. Measures of intensity posed the same challenges due to the varied nature of conversational speech. As can be seen in Figure 2, the distribution of average intensity measured over the duration of each fricative is very similar for /ð/ and /θ/ as well as /f/ and /v/, though /f/ has a slightly lower average intensity than /v/.

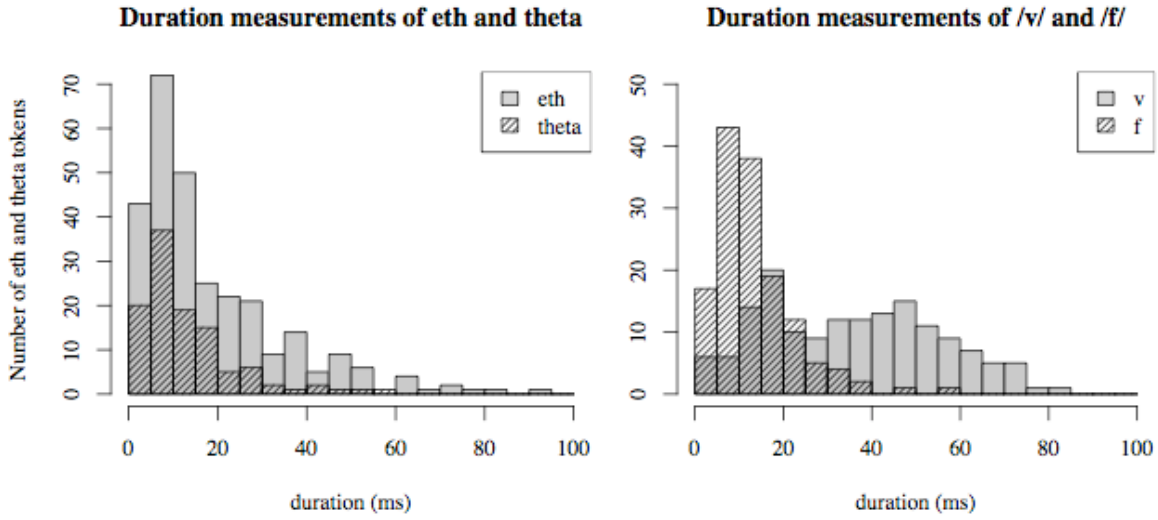

Figure 1: Duration of /ð/ and /θ/ (histogram on the left) and of /f/ and /v/ (on the right).

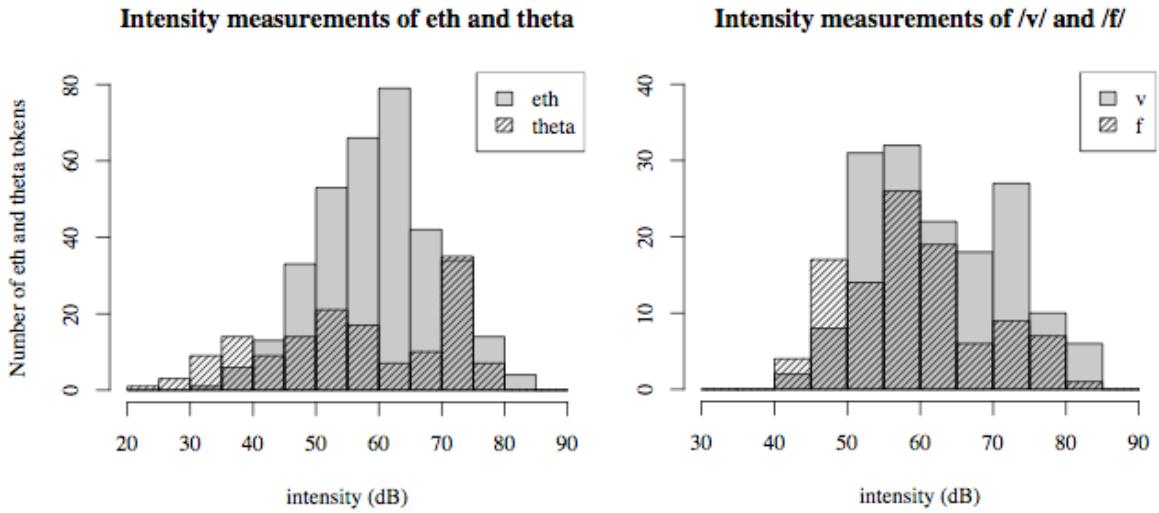

Figure 2: Average intensity of /ð/ and /θ/ (histogram on the left) and of /f/ and /v/ (right).

Because the overall duration measurements were more varied, and generally shorter than those in either Pirello et al. (1997) or Stevens et al. (1992), looking for 30 ms of voicing at either end of the fricative did not make sense. More than half of the fricatives were shorter than 30 ms overall. Because the tokens were not taken from CV sequences, but rather had very different environments, the measurement from Pirello et al. (1997) and Stevens et al. (1992), comparing the intensity of the first harmonic between the fricative and following vowel was not practical. Instead, the appearance of voice bar (periodicity of the waveform) was measured for each fricative interval, and taken as a percentage of the total duration of the fricative, as in Figure 3. The distribution for /f/ and /v/ appears bimodal, although there is a large amount of overlap. The majority of /f/ tokens are clustered below 20% voiced, while the majority of /v/ tokens are greater than 90% voiced. Again, /ð/ and /θ/ do not pattern exactly the same way as /f/ and /v/. There are two distributions, but not separated by phoneme. There is one distribution composed of both /ð/ and /θ/ tokens, clustered around 10-30% voiced, and another substantial cluster of 90-100% voiced tokens, composed primarily of /ð/ tokens, though including a few /θ/ tokens as well.

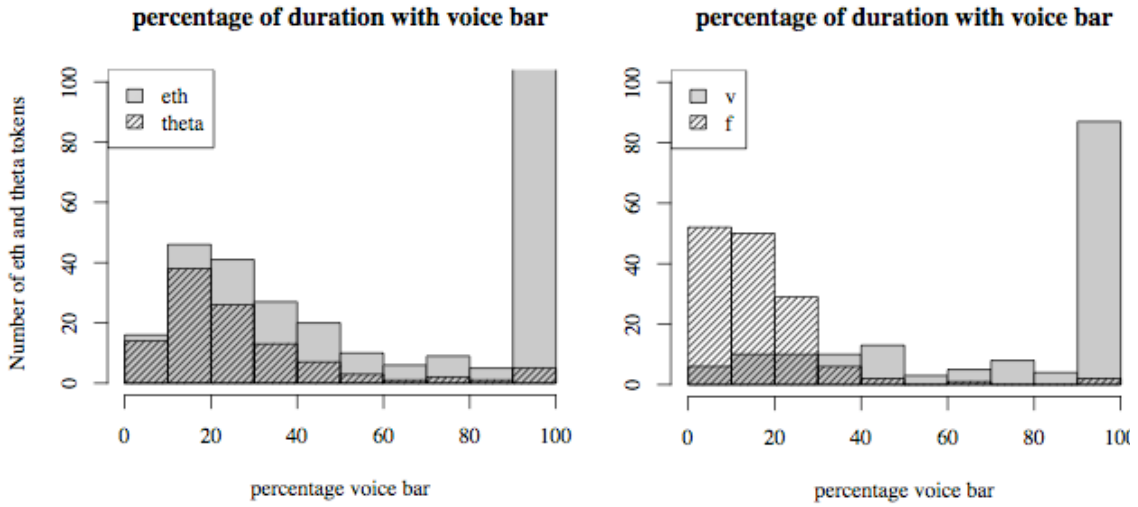

Figure 3: Percentage of the duration of the fricative that contained voice bar, with /θ/ and /ð/ on the left, and /f/ and /v/ on the right.

Yet somehow listeners are able to understand conversational speech. Because /ð/ and /θ/, and to a lesser extent, /f/ and /v/, maintain a similar pattern to their historical distribution, we might expect the voicing to pattern the same regardless of whether we separate the results by phoneme or by environment (that is, whether there is an adjacent voiceless sound, or if the segment is surrounded by voiced sounds). However, when we plot the same voice bar measurements as in Figure 3 by environment, as in Figure 4, a new pattern emerges. The dental fricatives develop a bimodal distribution, with predominantly voiceless sounds clustered on the left and predominantly voiced sounds on the right, distinguished not by phoneme, but by whether they were surrounded by voiced sounds or had a voiceless sound (including a pause) either preceding or following. The distribution of labio-dental fricatives shows a similar though somewhat murkier pattern, where this analysis appears to create more overlap, rather than reducing it.

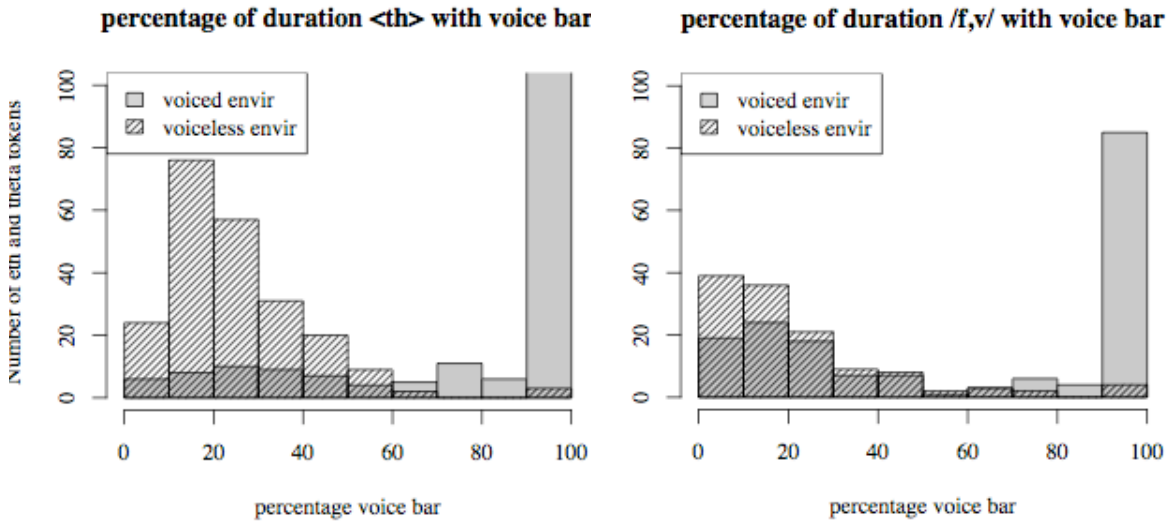

Figure 4: Percentage of the duration of the fricative that contained voice bar, with /θ/ and /ð/ combined on the left, separated by whether or not an adjacent segment was voiceless, and /f/ and /v/ combined on the right, also separated by voicing of adjacent segment.

Because the categories of phoneme and of environmental voicing largely overlap for these sounds, a partial correlation statistic was run on each of these data sets in order to determine how much variation was accounted for by phoneme, excluding that accounted for by environmental condition, and how much was accounted for by environmental condition, excluding that accounted for by phoneme, as in Table 2. Table 3 shows, by way of comparison, the results of a normal correlation of percent voiced by phoneme and by environment, separately.

Table 2: Partial correlation comparing variation accounted for by phoneme and environment, "partialing out" the variation accounted for by the other factor (environment and phoneme, respectively).

	r of voicing by ***phoneme***, with environment partialed out	**r** of voicing by ***environment***, with phoneme partialed out
dental fricative	0.4494	0.7661
labio-dental fricative	0.7186	0.4632

Table 3: Correlation of voicing and environment, and of voicing and phoneme.

	r of voicing by ***phoneme***, regardless of environment	**r** of voicing by ***environment***, regardless of phoneme
dental fricative	0.3666	0.7428
labio-dental fricative	0.7403	0.5157

The results of the correlation and partial regression show that phoneme category better accounts for the percentage of the duration that has voice bar in labio-dental fricatives, such that /v/ has generally more voicing, and /f/ less. But environment, that is, whether surrounding sounds are voiced, does a better job of accounting for percentage of voicing in dental fricatives, regardless of phoneme.

5. Discussion

The results of this investigation seem to indicate that the voicing contrast between the dental fricatives does not hold up well in conversational speech, even less well than /f/ and /v/, which are also highly variable. Duration and intensity are also not reliable measures of phonemic voicing in conversational speech. In fact, the very environments which conditioned voicing of the Old English dental fricative are more or less the same environments that predict voice bar in the modern dental fricatives. Even the "voiced" phonemes in function words can be produced as voiceless, although these are the same words that we assume were at least partially responsible for the phonologization of the voicing contrast. These results call into question whether the distinguishing feature between the two phonemes is actually voicing, or if there are some environments in which the contrast is neutralized. Because there is little to no competition in the interdental place of articulation, manner is free to vary, and indeed only 55% of the tokens of /θ/ and 29% of /ð/ that were analyzed in this experiment were realized as canonical fricatives; therefore, manner is also not a reliable indicator.

One thing that can be said for certain is that the variation of /ð/ occupies a much larger acoustic space, and encompasses that of /θ/, as illustrated in figures 1-3 for duration, intensity, and voicing, but also in place and manner, especially as it assimilates more readily to surrounding sounds, as in the greater number of tokens that were assimilated or deleted in this study. This raises some interesting questions. If the canonical forms are reported to be fairly confusable, but the conversational forms are even more variable and overlapping, how is it that we are able to distinguish them in speech? Or, because of the very low functional load carried by this contrast, do we even need to distinguish them? If voicing is predictable from environment, would it not be likely that listeners would have difficulty reconstructing an underlying form? Future studies are needed to explore perceptual aspects of and contrast or lack thereof in the dental fricatives.

References

Boersma, Paul & David Weenink. 2007. Praat: Doing phonetics by computer (Version 4.5.14) [Computer program]. Retrieved February 2007, from [http://www.praat.org/].

Bybee, Joan, & Paul Hopper. 2001. Editors. *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins.

Dalen, Arnold. 2002. "Sources of written and oral languages in the 19th century". In *The Nordic Languages: An international handbook of the history of the North Germanic languages.* Vol 2. Ed. by Oskar Bandle, Kurt Braunmuller, Ernst Hakon Jahr, Allan Karker, Hans-Peter Naumann, & Ulf Teleman. Berlin: DeGruyter. 1406-1418.

Hughes, Arthur, Peter Trudgill, & Dominic Watt. 2005. *English accents and dialects: An introduction to social and regional varieties in the British Isles*. 4th ed. London: Trans-Atlantic Publications, Inc.

Maddieson, Ian, & Kristin Precoda. 1990. UPSID-PC *The UCLA Phonological Segment Inventory Database.* (Data on the phonological systems of 451 languages, with programs to access it.) Accessed from [http://www.linguistics.ucla.edu/faciliti /sales/software.htm]

Pirello, Karen, Sheila E. Blumstein, & Kathleen Kurowski. 1997. "The Characteristics of Voicing in Syllable-Initial Fricatives in American English". *Journal of the Acoustical Society of America* 101.3754-3765.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume, & Eric Fosler-Lussier, comps. 2006. *Buckeye Corpus of Conversational Speech*. (1st release) [www.buckeyecorpus.osu.edu] Columbus, Ohio: Department of Psychology, Ohio State University (Distributor).

Polka, Linda, Connie Colantonio, & Megha Sundara. 2001. "A cross-language comparison of /d/ - /ð/ perception: Evidence for a new developmental pattern." *Journal of the Acoustical Society of America.* 109(5):2190-2201.

Raphael, Lawrence J. 1972. "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English." *Journal of the Acoustical Society of America.* 51:1296-1303.

Smith, Bridget J. 2007. "The seeds of sound change don't fall far from the tree". Unpublished ms. presented at *OSU Colloquium Fest*, Nov. 9, 2007.

Smith, Bridget J. 2009. "Dental fricatives and stops in Germanic." In Dufresne, Dupoux, & Vocaj (eds.), *Historical Linguistics 2007. CILT 308*. 19-36. Amsterdam/Philadelphia: John Benjamins.

Stevens, Kenneth N., Sheila E. Blumstein, Laura Glicksman, Martha Burton, & Kathleen Kurowski. 1992. "Acoustic and Perceptual Characteristics of Voicing in Fricatives and Fricative Clusters". *Journal of the Acoustical Society of America.* 91: 2979-3000.

Wells, John C. 1982. *Accents of English 2: The British Isles*. Cambridge University Press. 1982.

Wolfram, Walt. 1970. "Some illustrative features of Black English." Workshop on Language Differences. Coral Gables, FL, February 1970

Wolfram, Walt. 1974. *Sociolinguistic aspects of assimilation: Puerto Rican English in New York City.* Arlington, VA: Center for Applied Linguistics.

Appendix: Characteristics of manner of articulation

Criteria used for manner judgments presented in Table 1.

(A) Voiceless fricative: Characterized by aperiodic noise, sometimes forming a diamond or triangle shape in the waveform as intensity increases over the course of the fricative. Because /ð/ and /θ/ are very quiet, the aperiodic noise may maintain a very low intensity throughout. It sounds like a traditional <th> sound, as in *thing*. There may be one or more quiet bursts, but if the frication leading up to and following the burst are of a similar character, and the burst is quiet, and the segment sounds like a fricative, it is labeled as a fricative rather than a stop. Bursts may occur in fricatives as the oral air pressure builds up behind the constriction and is released. This can happen if the lungs are sending out a little too much air to fit through the narrowed constriction. Large bursts occur (and may be labeled as a stop) if the constriction is too narrow and the air pressure too high, so that the air gets really backed up and the constriction acts as a (leaky) closure. See section (C) on stops.

(B) Voiced fricative: Sounds like <th> in the word *then* or *breathe*. There should be frication in the higher frequencies, visible as aperiodic noise in the waveform, as well as regular periodicity indicating voicing for most of the duration. Sometimes the frication may be quiet, and only barely distinguishable from background noise, but there should be at least some fuzziness (aperiodic noise) in each period. Though there will be regular voicing, the waveform will have much less intensity than nearby vowels or approximants, and the spectrogram will not have clear formants, especially above F3.

(C) Stop: Noted where there is a (larger, usually audible) burst, with or without voicing, and with or without aspiration. Usually there will be some period of closure (with or without voicing) before the burst, followed by aspiration or frication. This will sound like it has a much more aggressive onset that that of a regular fricative. There may be little or no aspiration after the fricative (as in a short-lag or voiced stop) or there may be aspiration and/or frication as in an aspirated stop. (But if the frication is very strong and/or long, it may be labeled as an affricate.)

(D) Approximant: A dental fricative realized as an approximant may sound vaguely /l/ -like. Formants may be clear or faint, but they have noticeably reduced amplitude relative to surrounding vowels. There should be no frication noticeable in the waveform or spectrogram. Flaps look like approximants that have slightly shorter duration and reduced amplitude, and especially unclear or absent formants, but there are many tokens that are ambiguous. Flaps are also labeled as approximants because there is no good way of distinguishing between them.

(E) Vowel: A dental fricative may be realized as a vowel, usually a /ə/ or /□/ type sound. The waveform is (nearly) indistinguishable from surrounding vowel sounds, and there are very clear formants, also (nearly) indistinguishable from surrounding vowels. The difference between the vowel and approximant is one of amplitude. Generally, approximants have reduced amplitude relative to vowels, and fainter formants. If there is

no decrease in amplitude, it is a vowel. If no trace of the dental fricative can be heard, then it is marked as a deletion. If there is a change in the vowel quality corresponding to the percept of a dental fricative, then the segment is labeled as a vowel.

(F) Nasal: It has antiformants, reduced amplitude, and is often next to another nasal. The nasality is clear in the spectrogram, but is also audibly obvious. If the combined segment is as short as a single segment, the dental fricative may just be deleted. If the nasal is long, the dental fricative is assimilated and realized as a nasal.

(G) Lateral: While approximant realizations may sound somewhat /l/-like, segments marked as laterals are obvious /l/s. These usually occur as complete assimilations to neighboring segments. If the original /l/ plus dental fricative as /l/ looks longer than a single segment, the dental is realized as a lateral. If the combined segment is as short as a single segment, the dental fricative may just be deleted.

(H) Affricate: A combination of a stop+fricative

(I) Mixed bag: If the sound has two distinct manner realizations, such as fricative+stop or fricative+approximant, both are noted.

PROSODY OF FOCUS AND CONTRASTIVE TOPIC IN K'ICHE'

Murat Yasavul*
Ohio State University

Abstract

This paper discusses the findings of an experimental study about the prosodic encoding of focus and contrastive topic in K'iche'. The central question being addressed is whether prosody plays a role in distinguishing string-identical sentences where the pre-predicate expression can be interpreted as being focused or contrastively topicalized depending on context. I present a production experiment designed to identify whether such sentences differ in their prosodic properties as has been impressionistically suggested in the literature (Larsen 1988; Aissen 1992; Can Pixabaj & England 2011). The overall strategy of the experiment was to obtain naturally occurring data from native speakers of

*I am indebted to the speakers of K'iche' who participated in this study and to Raul Castro, María Hernandez Us, Adelina Chom Canil and Juana Pérez Gómez for their judgments about the data I present. I also would like to thank Cynthia Clopper, Judith Tonhauser, Craige Roberts, Carl Pollard, Kathryn Campbell-Kibler, Mike Phelan, Laura Wagner, the participants of the Prosody-Semantics seminar in Fall 2010-Spring 2011 and the Prosody Working Group at Ohio State for their help with this study and for many helpful discussions about the material presented here. Of course, the usual disclaimers apply. I also thank Raul Castro, Heather Dean and Victoriano Canil for facilitating my fieldwork in Guatemala. The fieldwork for this project is funded by the Department of Linguistics and the College of Arts and Humanities at The Ohio State University.

Murat Yasavul. Prosody of focus and contrastive topic in K'iche'. *OSUWPL* Volume 60, Spring 2013, pp. 129-160.

K'iche' by having them repeat target sentences they heard in conversations. The phonological analysis showed that content words in K'iche' have a rising pitch movement, a finding which is in line with Nielsen (2005). The acoustic analyses of several variables yielded a significant effect of condition only in the range of the F0 rise associated with focused and contrastively topicalized expressions. However, the difference across conditions is only ∼6 Hz which may not be perceivable by listeners.

1 Introduction

In K'iche', a Mayan language of Guatemala, sentences like (1) may have two different interpretations given appropriate context (throughout, I use **boldface** for that part of the example which is relevant to the discussion at hand)[1]:

(1) **A** **Raul** x-∅-war-ik.
CLF Raul CMP-A3-sleep-SS
a. 'Raul slept.'
b. 'As for Raul, he slept.'

(1-a) is obtained when the pre-predicate expression *A Raul* 'Raul' is *focused*, i.e. when it is an answer to the Question Under Discussion (Roberts 1996), as in (1′)[2]:

(1′) Context: *Who slept?*
A **Raul**$_F$ x-∅-war-ik.
CLF Raul CMP-A3-sleep-SS
'Raul slept.'

(1-b), however, is obtained when the same expression is interpreted as a *contrastive topic*, the denotation of a topical constituent in a contrastive context (Roberts 2012), as in (1″):

[1]Unless otherwise stated, all the data in this paper are from original fieldwork in Santa María Tzejá, Ixcán, El Quiché, Guatemala and Columbus, Ohio, USA. In the orthography, all symbols have their standard phonetic value except the following: ' = glottal stop, C' = glottalized consonant, VV = long vowel, ch = [ʧ], tz = [ʦ], x = [ʃ], and j = [x] or [x̱]. The following abbreviations are used in the glosses of the examples: A1(p), A2(p), A3(p) = absolutive first, second, third person singular (plural) marker; E1(p), E2(p), E3(p) = ergative first, second, third person singular (plural) marker; 2s(p).f = second person singular (plural) formal; AFF = affectionate; AG = agent focus; AGT = agentive; AP = antipassive; ASP = aspect; CLF = classifier; COM = comitative; CMP = completive; COMP = complementizer; DAT = dative; DEM = demonstrative; DET = determiner; EMPH = emphatic; ENC = enclitic; FOC = focus particle; GEN = genitive; INCMP = incompletive; INSTR = instrumental; INTS = intensifier; IV = terminal suffix for morphologically intransitive verb; MOV = movement; NEG = negative particle; P>I = intransitive derived from positional; PART = particle; PERF = perfect; PL = plural; POS = positional; PREP = preposition; SS = status suffix; TOP = topic marker.

[2]The subscripts $_F$ and $_{CT}$ in K'iche' sentences indicate focused and contrastively topicalized expressions, respectively.

(1″) Context: A: *Raul and Roberto didn't work last night. Roberto went out.*
B: *And Raul, what did he do?*
A: **A** **Raul**$_{CT}$ x-∅-war-ik.
CLF Raul CMP-A3-sleep-SS
'As for Raul, he slept.'

A common property of focus and contrastive topic in K'iche', whose basic word order is predicate-initial, is that focused expressions may and contrastively topicalized expressions must be realized in the pre-predicate position. Additionally, such constituents can co-occur before the predicate as in (2), in which case the focused expression, here *al Maria* 'Maria', follows the contrastively topicalized expression, here *a Raul* 'Raul' (compare (2-a) and (2-b)). This fact provides language internal evidence that these discourse functions are distinguished by speakers.

(2) Context: *I know that Roberto saw Juana yesterday, but who did Raul see?*

a. A Raul$_{CT}$ al Maria$_F$ x-∅-r-il-o.
CLF Raul CLF Maria CMP-A3-E3-see-SS
'As for Raul, he saw Maria.'

b. #Al Maria$_F$ a Raul$_{CT}$ x-∅-r-il-o.
CLF Maria CLF Raul CMP-A3-E3-see-SS
(intended reading) 'As for Raul, he saw Maria.'

Alongside the change in basic word order, certain types of focus in K'iche' can be expressed by other morpho-syntactic means but these are neither obligatory nor do they apply across-the-board (more on this below). Consequently, this raises the question as to whether string-identical sentences like (1′) and (1″) differ in their prosodic properties because of the difference in their meaning. A broader question of interest is whether there always is a relation between pragmatics and prosody, in other words, whether meaning differences like the ones above are always reflected in the prosodic structure of otherwise identical sentences. Indeed, as regards K'iche', previous studies claimed that sentences like (1′) and (1″) differ in their prosodic properties. In particular, the literature has discussed whether the pre-predicate expressions in such sentences are set off from the rest of the sentence by a pause or not. Thomas Larsen (1991, p.c. cited in Aissen 1992) suggested that topics[3] in K'iche' are not followed by a pause. On the other hand, Can Pixabaj & England (2011) claimed that topics in K'iche', whether contrastive or not, are followed by a pause whereas foci are not.

In this paper, I discuss findings from a production experiment designed to identify whether the difference in meaning between focus and contrastive topic corresponds to a difference in prosody, which would distinguish string-identical sentences like (1′) and (1″) in K'iche'. The overall strategy of the experiment was to obtain naturally occurring data from native speakers of K'iche' by having them repeat target sentences they heard in conversa-

[3]There is no indication as to whether Larsen distinguished more than one kind of topic in K'iche'.

tions. The experiment was designed so that in each conversation, only one interpretation of the target sentence would be felicitous.

The rest of the paper is organized as follows. In section 2, I give the relevant background on K'iche' morpho-syntax which is necessary to understand the details of how focus and contrastive topic are expressed. In section 3, I summarize the literature on focus and contrastive topic in K'iche' in detail. I also elaborate on the differences between the current study and the previous work by making explicit my assumptions about focus and contrastive topic. After motivating the research question, I provide an overview of the previous work on the prosody of focus and contrastive topic in several languages including K'iche'. Section 4 presents the details of the production experiment, the analyses and the results before I conclude in section 5.

2 Background on K'iche' morpho-syntax

K'iche' is a Mayan language spoken by over a million people in the central and western highlands of Guatemala (Richards 2003). It has an ergative-absolutive agreement system (Larsen 1988) which is preserved throughout changes in aspect and clause type (Pye 2001). The basic word order is VS in intransitive clauses and VOA in transitive clauses (Larsen 1988; Pye & Poz 1988; England 1991), where S stands for the single argument of an intransitive, A for the more agent-like argument of a transitive, and O for the more patient-like argument of a transitive verb (Dixon 1994). In (3) and (4), I start with two examples that illustrate intransitive clauses.

(3) x-∅-war ri achi.
CMP-A3-sleep DET man
'The man slept.'

(4) x-at-war-ik.
CMP-A2-sleep-SS
'You slept.'

In K'iche', there is no case-marking on noun phrases, e.g. *ri achi* 'the man' in (3), to identify grammatical relations or semantic roles; these are read off of the verbal complexes via the ergative and absolutive cross-reference markers given in Table 1.

The absolutive markers are used to cross-reference, i.e. register the number and person features of, the S argument of an intransitive verb and the O argument of a transitive verb. In an intransitive verbal complex, e.g. *x-∅-war* 'CMP-A3-sleep' in (3), the sole argument *ri achi* 'the man' is cross-referenced by the phonologically null, third person singular absolutive marker -∅- 'A3' preceding the verb root *war* 'sleep'. The absolutive marker is preceded by the aspect marker *x-* 'CMP'. In (4), where the argument of the verb is not real-

Ergative	Preconsonantal	Prevocalic
E1	*-in-*	*-inw-/-w-*
E2	*-aa-/-a-*	*-aw-*
E3	*-uu-/-u-*	*-r-*
E1p	*-qa-*	*-q-*
E2p	*-ii-*	*-iw-*
E3p	*-ki-*	*-k-*

Absolutive	
A1	*-in-*
A2	*-at-*
A3	*-∅-*
A1p	*-uj-/-oj-*
A2p	*-ix-*
A3p	*-e'-/-eb'-/-ee-*

Table 1: Ergative and absolutive agreement markers

ized, the verbal complex also carries the status suffix *-(i)k* 'SS' following the verb root. This marker is claimed to mark phrase-finality, in particular, the end of an intonational phrase Henderson (2012) and it is used for intransitive verbs in the incompletive and completive aspects.

The other set of markers in Table 1, namely the ergative markers, are used to cross-reference the A argument of a transitive verb as exemplified in (5):

(5) x-at-u-to'-o.
CMP-A2-E3-help-SS
'S/he helped you.'

In a transitive verbal complex, e.g. *x-at-u-to'-o* 'CMP-A2-E3-help-SS' in (5), the absolutive marker *-at-* 'A2', which marks the O argument of the verb, precedes the ergative marker *-u-* 'E3', which marks the A argument. The ergative marker, in turn, precedes the verb root *to'* 'help'. Similar to intransitive verbs, transitive verbs may carry phrase-final suffixes when they occur at the end of intonational phrases (Henderson 2012). For example, in (5) the verb root is followed by the status suffix *-o*[4].

Since K'iche' does not use overt marking on noun phrases, it is the absolutive marker *-∅-* 'A3' that identifies, say, *ri achi* 'the man' as the O argument in (6) and it is the ergative marker *-u-* 'E3' that identifies *ri achi* in (7) as the A argument (Trechsel 1993):

(6) x-∅-a-to' ri achi.
CMP-A3-E2-help DET man
'You helped the man.'

(7) x-at-u-to' ri achi.
CMP-A2-E3-help DET man
'The man helped you.'

[4]The form of the status suffix for transitive verbs can be *-u*, *-o* or *-j* depending on the derivational status of the stem (Trechsel 1993). The status suffixes simultaneously register (in)transitivity, aspect and, in the case of transitive verbs, the derivational status of the stem (Pye 2001).

Although the basic word order in K'iche' is VS/VOA, in texts it is relatively uncommon to find the A, O or the S arguments in post-predicate positions realized as pronominal arguments. Larsen (1987:40) claims that independent pronouns, which rarely appear in argument positions, are used in some cases to indicate "contrastive emphasis" or change of subject[5]. These pronouns, given in Table 2 below, are identical to absolutive markers except for the third person singular and plural.

1sg	*in*
2sg	*at*
3sg	*are'*
1pl	*oj*
2pl	*ix*
3pl	*e a're'/a're'/ke*

Table 2: Pronouns in K'iche'

In addition to these pronouns, K'iche' marks formality/politeness for second person by two morphemes[6]: (i) *la* 'you' (singular) and (ii) *alaq* 'you' (plural) which occur post-verbally (Trechsel 1993). In (8), *la* '2s.f' marks the formal second person singular ergative argument and *oj* 'A3' marks the absolutive argument. In (9), *alaq* '2p.f' marks the formal second person plural argument:

(8) x-oj-to' **la**.
CMP-A3-help 2s.f
'You (sg. formal) helped us.'

(9) x-pee **alaq**.
CMP-come 2p.f
'You (pl. formal) went.'

The two word orders I discussed in this section characterize basic, non-emphatic sentences, i.e. sentences which do not involve topicalization or focus, and in which pronominal arguments are usually dropped. After this basic description of the relevant morpho-syntactic properties of K'iche', I now turn to the main topic of the paper, namely how the two discourse functions focus and contrastive topic are expressed.

[5]In the discussion of focus and topic below, we will see that these pronouns can occupy the pre-predicate positions when they are focused or topicalized.

[6]The formal pronouns will become relevant later on in the discussion of agent focus marking.

3 Focus and contrastive topic in K'iche'

3.1 Previous literature on focus in K'iche'

A general claim about Mayan languages, dating back to Norman (1977), is that they are generally predicate-initial, but that there are also two special positions preceding the predicate that constituents can occupy for pragmatic purposes. The discourse functions that these constituents have, which are called *focus* and *topic*, govern the changes in the basic word order in K'iche' (Larsen 1988; England 1991). Norman (1977) claimed that focus and topic are structurally different in that focus occupies the pre-predicate position whereas topic occurs sentence initially.

Focus constructions in Mayan have been traditionally analyzed as involving a movement operation whereby the focused constituent is realized in the pre-predicate position and linked to a gap in the post-focal portion of the sentence (Larsen 1988; Aissen 1992; Trechsel 1993)[7]. In her seminal work on topic and focus in Mayan, Aissen (1992) claimed that focused constituents occupy the [Spec, I′] position and bind a co-indexed trace lower in the tree. The constituents occupying the focus position are generally understood to be semantically "prominent" in some sense (Larsen 1988) as reflected in the cleft translation into English in (10), which is the standard practice in the Mayan literature (Aissen 1992; Larsen 1988; Trechsel 1993; Can Pixabaj & England 2011):

(10) **Aree ri achi** x-∅-q'ab'ar-ik.
FOC DET man CMP-A3-get.drunk-SS
'It was the man who got drunk.' (Larsen 1988:503)

Aissen (1992:43), in particular, claims that focus in Mayan has the two characteristics associated with the interpretation of clefts: an existence presupposition and a uniqueness assertion. The following example from Aissen (1992:49) is taken from the middle of a text in Tzotzil, where one individual, walking along, meets another working in a field who utters (11-a) and the narrative continues with (11-b). According to Aissen, in (11-b) there is a presupposition to the effect that there was something that the man was planting and that the focused expression *chobtik* 'corn' is the unique entity that satisfies this presupposition:

(11) a. 'I'm planting. I'm planting stones, I'm planting trees',
b. Pero **chobtik** tztz'un un.
but corn he.plants ENC
'But it was corn he was planting.' (Aissen 1992:49)

[7]In fact, Mayanists have traditionally subsumed pre-predicate focus constructions, content questions and relative clauses under the heading of focus because they characterized these constructions by the obligatory presence of a constituent preceding the predicate, the obligatory gap in the post-focal portion of the sentence, a dependency between them and the use of agent focus form (Larsen 1988; Trechsel 1993).

Focus constructions in Mayan are further characterized by a special verb form called the *agent focus* form, a much discussed phenomenon in the context of focus (see e.g. Mondloch 1981; Larsen 1988; Trechsel 1993; Aissen 2011 for K'iche' and Dayley 1981; Aissen 1999; Stiebels 2006 for other Mayan languages). Agent focus can only be used with transitive verbs when the ergative argument of the verb is focused as in (12):

(12) Aree le achi x-∅-kuna-**n** le ixoq.
FOC DET man CMP-A3-cure-AG DET woman
'It was the man who cured the woman.' (Trechsel 1993:42)

The verbal complex in (12), *x-∅-kuna-n* 'CMP-A3-cure-AG', is in the agent focus form which is expressed by (i) the absence of an ergative marker *-u-* 'E3' on the verb, and (ii) the presence of the agent focus marker[8] *-n* 'AG' attached to the verb. In (12), both the agent and the patient are third person singular and it is indeterminate whether it is the agent or the patient that the absolutive marker agrees with. Yet, when there is an agent focus marker, the interpretation is always that the pre-predicate argument, which denotes the agent of the action, is focused. Larsen (1988) points out that the agent focus form can never be used in simple transitive clauses.

In a recent study on K'iche' texts, Can Pixabaj & England (2011) argue that there are two types of focus in K'iche'. The first type is what they call *contrastive focus* which "usually requires an explicit contrast" and which, they claim, operates like clefts in English (p.23). In (13), for instance, Can Pixabaj & England (2011:22) say that the focused expression "explicitly contrasts 'my parents' with 'me', identified negatively in the previous clause ('it wasn't I who saw')":

(13) Pero **aree r-in-taat** k-e-tzjo-**n**-ik.
but FOC DET-E1-father INCMP-A3p-recount-AG-SS
'...but it was my parents who recounted (it).' (Can Pixabaj & England 2011:22)

According to Can Pixabaj & England, this kind of focus requires the use of the focus particle *aree* 'FOC' with definite nominals[9] as well as the agent focus form of the verb when ergative arguments are focused as in (13). They also claim that this type of focus is not followed by a pause (p.21)[10].

[8]This marker comes in two forms: *-(V)w* for root transitive verbs, and *-n* for derived transitive verbs (Trechsel 1993).

[9]Regarding the definite-indefinite distinction in K'iche', Can Pixabaj & England say "[w]e consider those that have no article or possessor, or have only the indefinite article *jun* to be "indefinite", while we consider those that are accompanied by one of the definite articles *wa, le, ri* (with or without the indefinite article), are possessed, are accompanied by demonstratives, or are proper names to be "definite"". They also claim that *xow* 'only' can precede definite nominals in contrastive focus contexts but they do not provide examples.

[10]The source of the examples in this study is based on five texts with more than 1,800 clauses. The commas in the texts after expressions in the pre-predicate position are taken to indicate pauses, and the lack thereof as evidence that there are no pauses.

The second type of focus that Can Pixabaj & England (2011:23) identify is used to "present new information", "mention a participant for the first time" or "reintroduce information". This type of focus is not used for "explicit contrast of old information", nor does it require the use of the particle *aree* 'FOC' or the use of the agent focus form. Yet, similar to the first type focus, focused expressions of this type are not followed by a pause (p.23). The following example, the first sentence of a recording, illustrates "mentioning a participant for the first time" where "the speaker is identifying the person who will speak, from a pool of all who are present" (p.23):

(14) Chanim, **le don Santiago** k-∅-u-tzijoj cha-q-e jas le
now DET don Santiago INCMP-A3-E3-tell PREP-E1p-DAT what DET
u'istoria r-ech we jun tinamit Santa Lu's.
E3-history E3-POS DET one town Santa Lucía
'Now don Santiago will recount the history of the town of Santa Lucia.'
(Can Pixabaj & England 2011:24)

The following is an example where the focus "reintroduces a participant", *ri achi* 'the man', which "was spoken of about 50 clauses ago, using *rajawal* 'master'" (p.24):

(15) es ke **ri achi**, ri r-ajaw w-u'lew rii', ∅-k'o jun u-tajkil
it.is that DET man DET E3-master DET-land DEM A3-exist one E3-errand
aw-uuk'
E2-COM
'...it is that the man, he who is the master of this land, has an errand with you...'
(Can Pixabaj & England 2011:24)

Can Pixabaj & England do not explicitly provide the contexts in which these sentences are uttered. So, for instance, in (15) we do not know what the immediately previous context is and, therefore, we do not know whether *ri achi* 'the man' is focused or topicalized. Similarly, in (14), we do not know why *don Santiago* is necessarily identified as the person who will speak. This sentence may very well be an all focus sentence answering an implicit question like *What is going to happen now?* In fact, later on in the paper, Can Pixabaj & England (2011:26) note that this type of focus has the same *function* as a (non-contrastive) topic and the only difference between them is that the latter is followed by a comma in their textual data. All in all, given the lack of explicit contexts and definitions, it is hard to assess Can Pixabaj & England claims.

To summarize, we have seen that focus in K'iche', just as in other Mayan languages, can occur before the predicate. According to Aissen (1992), Larsen (1988) and Trechsel (1993) focus sentences are interpreted like clefts in English. According to Can Pixabaj & England (2011), however, K'iche' focus divides into two and only those sentences where the focused expression is preceded by *aree* 'FOC' are interpreted like clefts. Can Pixabaj

& England further claim that foci in K'iche', regardless of their type, are not followed by a pause. In the following section, I will elaborate on the assumptions I am making about focus and in doing so show how they can be applied to K'iche'. These assumptions are necessary to elaborate on my research question.

3.2 Background assumptions about focus

At an intuitive level, focus involves a way to mark "highlighted" or "emphasized" information in discourse. This seems to have been the general approach in the Mayan community in terms of its characterization of what is meant by the term focus. Despite ample discussion of this phenomenon in the literature along similar lines, the present study makes different assumptions about focus and how focus is expressed in K'iche'. Part of the reason for this departure from the common assumptions is empirical in that the generalizations made in the literature do not hold up against the data that I collected, which I illustrate below. Yet, the main motivation for a different characterization of focus in K'iche' is to situate it in the broader semantic-pragmatic literature and to have a principled characterization of focus that makes predictions. It might turn out that these assumptions need to be revised but the advantage of the framework that I will summarize below is that it gives us working definitions that we can test. It is not always clear what is meant by "new information", "emphasis" or "reintroducing a participant" etc. and without explicit definitions of these discourse functions, it is hard to come up with adequate analysis of the pragmatic phenomena that are under discussion.

A more principled characterization of focus is to consider it as answering an explicit or implicit question (Jackendoff 1972; Roberts 1996), which Kadmon (2001) claims to be the most basic and crucial intuition about focus. From this point of view, in a constituent question-answer pair, the phrase corresponding to the question-word is focused. So, for instance, in English the sentence *Michael ate tortillas*, with prosodic prominence on *Michael*, in particular a H* accent followed by a L-L% boundary tone[11] (Jackendoff 1972; Büring 2003) can constitute a felicitous answer to *Who ate tortillas?* (16) but not to, say, *What did Michael eat?* (17):

(16) Context: *Who ate tortillas?*
MICHAEL ate tortillas.
H* L-L%

[11]The letters L and H are used in the Autosegmental-Metrical (AM) Framework (Pierrehumbert 1980), which is a framework for intonational analysis. In the AM theory, the prosodic grouping and prominence relations are represented by distinctive pitch events, transcribed by a sequence of Low (L) and High (H) tones. or combinations thereof. The tones are marked with diacritics indicating their intonational function. There are *pitch accents* that mark prominence and *boundary tones* that mark the edges of prosodic boundaries. A star (*) on a pitch-accent indicates that it is associated with a stressed syllable. The % sign on a tone indicates a prosodic boundary.

(17) Context: *What did Michael eat?*
#MICHAEL ate tortillas.
H* L-L%

In (16), where the Question Under Discussion (QUD) is *Who ate tortillas?*, *Michael* corresponds to *who* whereas the rest of the sentence, *ate tortillas*, is congruent to the QUD in the sense that abstracting on the *wh*-word in the question yields the property $\lambda_x.x$ ate tortillas which is also the denotation of the rest of the sentence. Sometimes this partitioning with respect to a QUD is termed as the *Theme/Rheme* distinction where *Rheme* denotes the focus and *Theme* denotes the part of the sentence congruent to the QUD (Roberts 2012). A QUD is a semantic question, i.e. a set of propositions, that corresponds to the current discourse topic (Roberts 1996:93). It may be the denotation of an actual question that is asked as in (16) above or may be implicit in the discourse (Roberts 1998). As the examples above illustrate, focus presupposes that there is such a QUD, a presupposition which, together with contextual clues, enables the addressee to reconstruct, or *retrieve*, the QUD (Roberts 1996).

A related and widely-held view about focus is that it evokes alternatives in discourse (Rooth 1992). According to Rooth's analysis of focus interpretation, prosodic prominence on *Michael* in (16) evokes alternatives such as *Robert*, *Jane*, *Peter*, etc. with which one constructs a set of propositions of the form, x ate tortillas, for the original sentence where x ranges over possible alternatives drawn from a contextually restricted set E. This set of alternatives that focus evokes helps determine an additional semantic value for an utterance, which Rooth calls *the focus semantic value*. In other words, the focus semantic value of a focused expression α, denoted by $[\![\alpha]\!]^{\mathrm{f}}$, is obtained by making a substitution in the position corresponding to the focused expression in the sentence. To illustrate, the focus semantic value of (16) is given in (18). The ordinary semantic value can be drawn from the focus semantic value as the former is always an element of the latter (Rooth 1992:76). Crucially, the focus semantic value of (16) is the same set we obtain by abstracting on the *wh*-word in the question in (16), hence the question-answer congruence (Roberts 1996).

(18) $[\![[\text{MICHAEL ate tortillas}]]\!]^{\mathrm{f}}=\{\mathbf{ate}(x, \mathbf{tortillas})|x \in E\}$

So far, I have illustrated the question-answer congruence with English examples where focus is marked prosodically. Yet, Roberts (1996) points out that the prosodic realization of focus is not universally assumed by those working on the semantics of focus. This means that focus may involve non-prosodic means and, in fact, many languages use cleft-like structures, marked word order or special morphemes to indicate focus in addition to intonational marking (Büring 2011). Therefore, the common core of focus is the observation that it evokes alternatives and that it is intuitively linked to question-answer congruence irrespective of the actual means of realizing focus (Roberts 1996; Rooth 1996).

In the present study, I follow the line of thinking summarized above and characterize

focus in K'iche' as follows: (i) a focused expression can occur before the predicate and (ii) its meaning will yield an answer to the QUD when the meaning that is congruent to the QUD applies to it. In other words, an answer to the QUD, say in (16), is obtained by applying the property $\lambda_x.\ x$ ate tortillas to the meaning of the focused expression. The question-answer congruence, the defining characteristic of focus, can be shown in K'iche' as follows. Consider the examples in (19) and (20): (19-a)/(20-b) is a felicitous answer in (19) but not in (20), and (20-a)/(19-b) is a felicitous answer in (20) but not in (19):

(19) Context: *Who helped you?*

a. **A** **Raul**$_F$ x-in-u-to'-o.
CLF Raul CMP-A1-E3-help-SS
'Raul helped me.'

b. #**In**$_F$ x-in-u-to'-o.
I CMP-A1-E3-help-SS
'He helped me.'

(20) Context: *Who did Raul help?*

a. **In**$_F$ x-in-u-to'-o.
I CMP-A1-E3-help-SS
'He helped me.'

b. #**A** **Raul**$_F$ x-in-u-to'-o.
CLF Raul CMP-A1-E3-help-SS
'Raul helped me.'

I will end this section by discussing two properties of the focus stimuli that I used in the experiment. Recall that the research question of the present study builds on the observation that focus and contrastive topic sentences can be string-identical. In order for this to hold, the focus sentences should not carry any special focus marking except for the change in word order because contrastive topics, as we will see below, occur before the predicate with no additional morpho-syntactic marking. Consequently, none of the focus stimuli had the focus particle *aree* 'FOC' in them and, furthermore, when ergative arguments were focused, the agent focus marker wasn't used.

Although it is widely discussed as a concomitant of focusing ergative arguments, the agent focus marker was not obligatory for my informants and they were not making use of this form very often in elicitation sessions. Larsen (1988:505) also reports that using agent focus is optional even when its use is permissible. In any case, there are restrictions regarding the use of agent focus. For instance, at least one of the arguments of the verb has to be third person or second person formal for the use of agent focus to be felicitous:

(21) *__In__ x-at-ch'ay-__ow__-ik.
I CMP-A2-hit-AG-SS
(intended reading) 'I hit you.'

In order to focus the agent NP in (21) one can: (i) use the active voice as in (22), or (ii) demote the patient NP and use the oblique phrase *aw-e* 'E2-GEN' as in (23):

(22) **In** x-at-in-ch'ay-o.
I CMP-A2-E1-hit-SS
'I hit you.'

(23) **In** x-in-ch'ay-**ow** **aw-e**.
I CMP-A2-hit-AG E2-GEN
'I hit you.'

In summary, the agent focus form cannot always be used. Even when it is applicable, there are either restrictions on its use or it alternates with the active form of the verb. Moreover, there is no counterpart of agent focus for intransitive verbs or for cases where the O-argument of a transitive verb is focused. When these arguments are focused, the active form of the verb is used (Larsen 1988; Trechsel 1993). This shows that in general foci can be marked by only a change in word order just as topics. In the next section, I turn to the discussion of topics in K'iche'.

3.3 Previous literature on topics in K'iche'

The second discourse function that can be expressed by a pre-predicate expression in K'iche' (and in Mayan languages in general) is called *topic*. The topic of a sentence is defined to be the constituent that indicates what the sentence is about (Aissen 1992; Roberts 2012) and in this sense a topicalized expression is an entity to which our attention is drawn (Aissen 1992; Roberts 2012). Below is an example in K'iche' where the topicalized expression *Ri ulew* 'the earth' precedes the predicate:

(24) Context: *Tell me something about the earth.*
Ri **ulew** k-∅-b'in chi-r-ij ri q'iij.
DET earth INCMP-∅-walk PREP-E3-around DET sun
'The earth revolves around the sun.'

Aissen (1992) distinguishes between two kinds of topics in Mayan languages: (i) *external topics* and (ii) *internal topics* and argues that these topics behave differently both structurally and pragmatically. The following example from Tzotzil illustrates external topics. Aissen points out that the first line in (25) introduces two discourse participants, the second line turns attention to one of them, namely *a ti vinik-e* 'the husband' and asserts something about him, and the third does the same for the other participant, here *a ti antz-e* 'the wife'. Both of the topicalized expressions are preceded by the topic marker *a* 'TOP'. They are also usually accompanied by a definite determiner *ti* 'DET' and an enclitic *e* 'ENC'.

(25) a. There was a man and a woman, newlyweds.
b. **a ti vinik-e** ta-xlok' ech'el, ta-tbat ta-xxanav.
TOP DET man-ENC exists away goes travels
'The husband leaves, he goes, he travels.'
c. **a ti antz-e** jun-yo'on ta-xkom
TOP DET woman-ENC happily stays
'The wife stays at home happily...' (Aissen 1992:49)

Aissen refers to external topics as new or shifted topics: once a participant is topicalized in this way, it is not referred to again by an overt nominal unless the topic shifts to another participant (p.51). Structurally, external topics occupy a position outside the clause, as a sister of the CP, and are base-generated. There is no requirement that they bind a coreferential pronoun lower in the clause[12]. Their structure, therefore, resembles that of left-dislocation (p.48) where the topic is prefixed to a fully well-formed root CP so long as the CP is about the topic. Aissen also makes a claim about the prosody of such topics and says they are followed by a pause, which, in her theory, follows from the syntactic structure (p.76).

The second kind of topic Aissen identifies, namely internal topics, involves discourse participants which are already identified as topic and can occur in the pre-predicate position. The following piece of discourse in Tz'utujil provides an example of such topics. The texts starts '[a] long time ago there was a man whose daughter was in a dance' and Aissen claims that (26-a) introduces *rme'al* 'his.daughter' as a new topic, marked by the particle *ka'(ar)* 'PART', and this same topic is referred to again by an overt nominal in the following sentence in (26-b):

(26) a. **Ja k'a rme'al** x-u-koj pa xajoj xin Tukun.
the PART his.daughter ASP-E3-enter in dance of Tecun
'He entered his daughter in the dance of Tecun.'
b. y **ja rme'al** x-ok-i Malincha.
and the his.daughter ASP-play-IV Malincha
'and the daughter played the part of the Malincha.' (Aissen 1992:74-75)

Such NPs can occur in the topic position although their referent has already been established. Structurally, these topics occupy the [Spec, C′] position, and like foci, bind a co-indexed trace lower in the clause. Furthermore, these topics are not separated from the following clause by a pause. As regards K'iche', Thomas Larsen (p.c. 1991, as cited in Aissen 1992) suggested that topics in K'iche' have the function of external topics in terms of their meaning but are associated with the syntax of internal topics, i.e. there is no pause after them.

[12]In Jakaltek, such topics may bind overt pronouns in the CP and yet in Tzotzil we don't find these pronouns as the language is pro-drop (Aissen 1992:69).

Building on Aissen's work, Can Pixabaj & England (2011) argue that there are two types of topics in K'iche'. Their characterization of topics is structural in that they are interested in "defining structurally the preverbal positions that can be filled by noun phrases". According to their characterization, the first type of topic occurs in "the first position" (sentence-initial; MY) preceding the verb and has no "special" marker such as *aree* 'FOC' or "special" verb form (agent focus form; MY) when it is the subject of a transitive verb (p.19). An example is (27) where Can Pixabaj & England claim that the hunter "was introduced in the previous clause and is here established as the local topic and continues as such for three more clauses, with only anaphoric reference" (p.20):

(27) **Ri k'aq-an-eel**, iii b'yeen ∅-u-b'an-om k'ax ch-k-e
DET hunt-AP-AGT eh INTS A3-E3-do-PERF bad PREP-A3p-DAT
s-taq-a'waj-iib'.
AFF-PL-animal-PL
'The hunter had done much damage to the animals.'
(Can Pixabaj & England 2011:20)

As with the examples of the different kinds of focus above, Can Pixabaj & England do not provide the context in which this sentence is uttered. Therefore, we do not know how the expression *ri k'aqaneel* 'the hunter' was introduced and whether it occurred in the pre- or the post-predicate position nor do we know whether it was focused or topicalized in the previous clause.

The second type of topic that Can Pixabaj & England (2011) identify is called *contrastive topic* which combines the functions of topic and focus "in the context of changing the topic and at the same time contrasting it with the previous topic" (p.24). According to Can Pixabaj & England, such topics can be preceded by the phrase *aree k'u*[13]. Unlike "contrastive focus", however, there is no "special" verb form (agent focus form; MY) that can be used with this construction. Furthermore, the nominal which is contrastively topicalized is followed by a pause. An example that Can Pixabaj & England (2011) provide is (28) where they say that in clauses before this example "the topic was the hunter, now it is the master of the mountain where he went to hunt" (p.24):

(28) Tonse are k'u **ri r-ajaw-al** u-winaq-il ri' ri jyub',
well EMPH PART DET E3-master-ABST E3-person-ABST DEM DET ill
jawi r-qas -k-∅-e'-k'aqa-n-a wi, x-∅-tak'-i'
where DET-always INCMP-A3-MOV-hunt-AG-SS EMPH CMP-A3-standing-P>I
r-oyowaal.
E3-anger
'Well, on the other hand the master of the hill, where he always went to hunt, got mad.' (Can Pixabaj & England 2011:25)

[13]A similar claim made by López Ixcoy (1997) is that the particle *aree* 'FOC' itself precedes contrastively topicalized expressions.

Given that Can Pixabaj & England take topics to occur only before the predicate, it is safe to assume that *the hunter* occurs before the predicate in the clause preceding (28). Yet, we do not know why *the hunter* was the topic rather than the focus because Can Pixabaj & England do not provide the context in which the sentence is uttered. This lack of contextual evidence makes it hard to determine the discourse status of the pre-predicate expressions in the examples they present.

To summarize, we have seen that in Mayan topicalized expressions occur before the predicate and generally two kinds of topic are distinguished. Aissen's external topics and Can Pixabaj & England's contrastive and non-contrastive topics are all separated from the post-topical portion of the sentence by a pause. In the next section, I summarize the assumptions I am making about topics in K'iche' and point out the differences between the previous literature and the present study.

3.4 Background assumptions about topics

In this section, I will present the assumptions I am making about topics in K'iche' and the kinds of topical constituents that are realized before the predicate. The first kind of topic, which can be realized in the pre-predicate position, indicates what the sentence is about (29):

(29) Context: *What happened to Raul?*
A **Raul** x-∅-tzaq-ik.
CLF Raul CMP-A3-fall-SS
'Raul fell.'

The other kind of topical constituent, which, as far as my data suggest, is always realized in the pre-predicate position, behaves as a contrastive topic, i.e. the denotation of a topical constituent in a contrastive context (Roberts 2012). A contrastive topic, alongside being a topic, also implies that there is another question about a different topic. Put differently, it implies that there are other entities having the same type as the contrastively topicalized expression and that we are going through a list, so to speak, and answering the QUD with respect to the entity at hand. Consider the example below where, when the topic changes from *Raul* to *Roberto*, the new sentence answers the question with respect to *Roberto*:

(30) Context: A: *Raul and Roberto are farmers. Last year, Raul sowed corn.*
B: *And Roberto, what did he sow?*
A: **A** **Roberto**$_{CT}$ x-∅-u-tik kinaq'.
CLF Roberto CMP-A3-E3-sow beans
'As for Roberto, he sowed beans.'

Contrary to what Can Pixabaj & England (2011) and López Ixcoy (1997) claim, my consultants did not accept neither the marker *aree k'u* nor the marker *aree* with contrastively topicalized expressions. However, these markers were acceptable for them when the pre-predicate expression was focused. Consequently, contrastively topicalized expressions in my data do not carry the markers *aree* or *aree k'u* and, therefore, can be string-identical to focus sentences without *aree* and without agent focus marking.

3.5 Previous studies on the prosody of focus and contrastive topic

So far, I have shown how focus and contrastive topic are expressed in K'iche' and how a sentence with a pre-predicate focus can be string-identical to a sentence with a contrastive topic. I have also noted that the literature on K'iche' has discussed whether the focused or contrastively topicalized expression is set off from the rest of the sentence by a pause. There have been two opposite claims with respect to this issue: (i) Thomas Larsen (1991, p.c. cited in Aissen 1992) claims that K'iche' topics function as external topics in the sense of Aissen (1992) but are not followed by a pause, and (ii) Can Pixabaj & England (2011) claim that topics in K'iche' are followed by a pause regardless of their type whereas foci are not followed by a pause regardless of their type.

Whether there is a pause or not following pre-predicate expressions is one potential prosodic cue that is of interest to the present study. On the other hand, work on other languages has suggested that there are other prosodic cues associated with focus and contrastive topic which need to be taken into account in a thorough study of the prosody of focus and contrastive topic. In the following sections, I briefly summarize the cross-linguistic findings about the prosodic encoding of focus and contrastive topic. I first start with a summary of the work in languages other than K'iche' and then turn to the details about K'iche'.

3.5.1 Previous studies on other languages

It has been shown that prosodic prominence on a focused expression can be indicated by various phonological and phonetic means. In English, for example, it has been claimed that focus is primarily marked by a pitch accent, in particular by a H* pitch accent followed by a L-L% boundary tone (Jackendoff 1972; Büring 2003). In fact, it is argued that this intonational contour distinguishes focus from contrastive topic in English as the latter is marked by a L+H* pitch accent followed by a L-H% boundary tone (*ibid.*)[14]. In general, accenting has been taken as the primary source of prosodic prominence marking, at least for English (Rooth 1992; Kadmon 2001; Féry & Samek-Ladovici 2006).

For languages other than English, research has shown that prosodic prominence on a focused expression may be realized through a variety of phonetic and phonological

[14]The accents marking focus and contrastive topic have also been called A and B accent, respectively (Jackendoff 1972), and fall and fall-rise accent, respectively (Büring 2003).

means. For example, in some languages, e.g. Italian (Grice *et al.* 2005) and Spanish (Face 2002), different pitch accents are used to indicate focused expressions. Yet, in some other languages, e.g. Korean (Jun 2005) and Japanese (Venditti *et al.* 2008), prosodic prominence is realized through phrasing, namely by placing a prosodic phrase boundary before or after the focused expression to indicate prominence. In these languages dephrasing can be used to mark expressions as less prominent, which is similar to the use of deaccenting in English. These various phonological properties show that, cross-linguistically, different means are available to indicate prosodic prominence, e.g. accenting, phrasing.

Alongside these phonological means, many languages indicate prosodic prominence through phonetic means. For example, focused expressions in English are typically longer in duration (Cooper *et al.* 1985) and have an expanded pitch range compared to non-focused expressions (Eady *et al.* 1986). Similarly, in Mandarin, focused expressions have an increased pitch range and the pitch range of the post-focal expressions is compressed (Wang & Yu 2011). Another phonetic cue to prosodic prominence involves the alignment of the pitch accent peak. In Spanish, the alignment is earlier (Face 2001) whereas in German it is later on a focused expression (Braun 2006) compared to a non-focused expression.

As regards the prosody of contrastive topic, research has shown that such expressions have a particular prosodic structure, too. I have already noted above that contrastive topics in English are marked by a L+H* pitch accent followed by a L-H% boundary tone. In German, contrastively topicalized expressions carry a late-rising pitch accent and are prosodically separated from the main clause by a prosodic boundary (Féry 2006). In Mandarin, topics raise the initial pitch range but there is no prosodic correlate of contrastiveness of topics (Wang & Yu 2011).

In sum, prosodic effects of focus and contrastive topic can be indicated through both categorical phonological means and continuous phonetic means. An adequate study of the prosodic reflexes of discourse functions like focus and contrastive topic should take such means into account in the analysis.

3.5.2 Previous studies on K'iche'

Although the phonology of K'iche' is well-described (Mondloch 1978; López Ixcoy 1997; Larsen 1988), there are not many studies dedicated to its prosodic structure. Nevertheless, there have been some claims about the prosody of focus that I will present in this section.

A study devoted to a preliminary prosodic description of K'iche' is Nielsen (2005). Nielsen's work is different from all of the other work on K'iche' that makes claims about prosodic structure in that it involves intonational analyses of utterances from a native speaker rather than impressionistic claims or text analysis. In her study, Nielsen found that K'iche' has stress driven pitch accent and L+H* is the default pitch accent on content words. This finding is in line with the previous literature which claimed that K'iche' has word-final stress (Larsen 1988). Nielsen also described K'iche' as an accentual phrase

language where prosodic domains which may be slightly larger than a word, namely *accentual phrases*, are marked by a tone. According to Nielsen, the default L+H* accent on the prominent syllable of a content word also marks the boundary of an accentual phrase. Alongside these findings about the general prosodic structure of K'iche', Nielsen found upstepped pitch accents associated with focused expressions where the L tone of the L+H* associated with the focused expression starts higher than the previous L.

The other claims about the prosody of focus are related to the interaction between focus and negation. It has been traditionally claimed that negation in K'iche' is indicated by the negative particle *man*[15] before the predicate and the so-called irrealis particle *ta(j)*[16] after the predicate, with the form of *ta(j)* changing depending on where it occurs (Larsen 1988; López Ixcoy 1997; Can Pixabaj 2010; Henderson 2012). Henderson (2012) claims that the distribution of *taj* is the same as the status suffixes *-(i)k* and *-o*, i.e. it occurs at the end of intonational phrases[17]. Examples (31-b) and (32-b) below, which are the negated versions of (31-a) and (32-a), respectively, illustrate this variable pattern. In (31-b), *ta(j)* occurs at the end of an intonational phrase and is realized as *taj* whereas in (32-b) it assumes its non-phrase-final form and is realized as *ta* (Larsen 1988; Henderson 2012):

(31) a. X-∅-war-ik.
CMP-A3-sleep-SS
'S/he slept.'

b. **Man** x-∅-war **taj**.
NEG CMP-A3-sleep NEG
'S/he didn't sleep.'

(32) a. X-∅-inw-il ri achi.
CMP-A3-E1-see DET man
'I saw the man.'

b. **Man** x-∅-inw-il **ta** ri achi.
NEG CMP-A3-E1-see NEG DET man
'I didn't see the man.'

Larsen (1987:51) claims that when focus constructions are negated, the negation

[15]It has been reported that the negative particle *man* exhibits dialectal variation. In some dialects it is *man*, in some dialects it is *ma* and in yet others it is *na* (Larsen 1988; Henderson 2012).

[16]This particle has been traditionally glossed as an irrealis particle in K'iche' and it does have an irrealis meaning when it is used in counterfactual constructions Larsen (1988). However, it can be used without *man* in a negated sentence because, as Larsen points out, in many dialects of modern K'iche', the negative particle *man* is optional. In the speech of all but one of the consultants that I worked with, *man* is almost always omitted and only the so-called irrealis particle *ta(j)* is used. I, therefore, follow (Pye 2001) and treat *ta(j)* as a negation particle and gloss it as NEG in negated sentences.

[17]In the speech of my consultants, the non-phrase-final form *ta* is always realized in a reduced form as [t] cliticized to the preceding word. See Romero (2012) for a similar observation about the phonological realization of this particle.

particles are placed around the focused expression and, in particular, in the negated (33-b), the negation particle assumes its phrase-final form *taj* (translations are Larsen's):

(33) a. Are' x-∅-ch'ay-ow ri achi.
he CMP-A3-hit-AG DET man
'He was the one who hit the man.'
b. **Man** are' **taj/*ta** x-∅-ch'ay-ow ri achi.
NEG he NEG CMP-A3-hit-AG DET man
'He was not the one who hit the man.' Larsen (1987:51)

A conclusion that Larsen draws by comparing (33) to (34), where *taj* occurs before a clause boundary, is that since the focused expression in (33) is followed by the phrase-final form *taj*, there is a clause boundary immediately before the verbal complex showing that the focused constituent is separated from the post-focal material[18].

(34) Le achi **ma** x-∅-uu-chooma-j **taj** chi x-in-aa-ch'ay-o.
DET man NEG CMP-A3-E3-think-SS NEG COMP CMP-A1-E2-hit-SS
'The man didn't think that you hit me.' Larsen (1987:50)

However, Henderson (2012) claims that focused expressions form a phonological phrase in K'iche' and not an intonational phrase (pp.19-20). Therefore, the focused constituent cannot be followed by the phrase final form *taj* but rather by *ta*, the non-phrase final form of the negation particle:

(35) **Man** are' **ta(*j)** x-∅-r-il-o.
NEG s/he NEG CMP-A3-E3-see-SS
'S/HE didn't see him/her. Henderson (2012:19)

Before going into the details of the experiment, I will mention some relevant points about the claims we have seen so far for the study at hand. For instance, if focused expressions in the data are followed by intonational phrase boundaries, then that will provide counterevidence for Henderson's (2012) claim that focused constituents do not form intonational phrases. If, on the other hand, the pre-predicate expressions are not followed by pauses then that would provide counter-evidence to the claim put forth in Can Pixabaj & England (2011) that topics in K'iche' are followed by a pause. Since the focused expressions are not preceded by any other expression in the experimental stimuli, it is not possible to test whether Nielsen's (2005) claim about up-stepped pitch accents holds true for my data. Yet, it is possible to see if her description of the prosodic structure of K'iche' is reflected in the data I collected. These and the claims about the prosodic reflexes of focus and contrastive

[18]Larsen's claim is not necessarily a prosodic one as he conceives of the boundary as a syntactic clause boundary (p.51).

topic in other languages will be taken into account in the prosodic analyses of the data.

4 The experiment

The discussions in the previous sections provide support for the claim that K'iche' allows string-identical sentences to have different interpretations as in (36) repeated here from section 1:

(36) **A Raul** x-∅-war-ik.
CLF Raul CMP-A3-sleep-SS
a. 'RAUL slept.'
b. 'As for Raul, he slept.'

This raises the question as to whether such sentences differ in their prosodic properties. In order to answer this question, I designed and carried out a production experiment with native speakers of K'iche' which aimed to obtain naturally occurring data. In a nutshell, the experiment involved participants listening to conversations accompanied by visual stimuli. The last sentence of each conversation was a target sentence where the pre-predicate expression was interpreted either as a focus or a contrastive topic depending on context. The task for each participant was to utter the target sentence as an answer to a question that is part of the conversation s/he heard. The following sections lay out the details of this experiment.

4.1 The participants

The experiment was carried out with 6 (4F, 2M) native speakers of the Joyabaj dialect of K'iche' in Santa María Tzejá, Ixcán, Guatemala in the summer of 2011. All of the participants were bilingual in K'iche' and Spanish and non-literate in K'iche'. They did not report any hearing, speech or visual impairments. The speakers were paid for their participation in the study.

4.2 Methods

4.2.1 The stimuli

The stimuli of the experiment consisted of 32 context-target utterance sequences (16 contrastive topic, 16 focus) and 19 fillers consisting of similarly constructed discourses. An example discourse for a focus sentence, which consists of a question-answer pair, is given in (37):

(37) A: Chin x-∅-u-b'an le wa?
who CMP-A3-E3-make DET tortillas
'Who made the tortillas?'
B: **Al Maria**$_F$ x-∅-u-b'an le wa.
CLF Maria CMP-A3-E3-make DET tortillas
'Maria made the tortillas.'

The corresponding discourse where the pre-predicate expression is contrastively topicalized is given in (38). Here, speaker B introduces two discourse participants, *Maria* and *Manuela*, and says something about *Manuela*. Speaker A then asks about *Maria*, and B's answer to that question is the target sentence which is string-identical to the one in (37). Note that in the case of contrastive topic, the stimulus is not just a question-answer sequence but rather has an additional sentence which introduces two discourse participants before the question is asked.

(38) B: Al Maria r-ichbil al Manuela x-∅-ki-b'an rikil.
CLF Maria E3-and CLF Manuela CMP-A3-E3p-make dinner
'Maria and Manuela made dinner.'
Al Manuela x-∅-u-b'an le kinaq'.
CLF Manuela CMP-A3-E3-make the beans.
'Manuela made the beans.'
A: Al Maria, su x-∅-u-b'an-o?
CLF Maria what CMP-A3-E3-make-SS
'Maria, what did she do?'
B: **Al Maria**$_{CT}$ x-∅-u-b'an le wa.
CLF Maria CMP-A3-E3-make DET tortillas
'As for Maria, she made the tortillas.'

As in the examples above, each discourse was constructed in a way to make only one interpretation of the target sentence possible. The target utterance was always the last sentence of a given discourse and was always an answer to a question eliciting a focused or contrastively topicalized expression. All of the pre-predicate expressions in the target sentences were proper names with penultimate stress. This enabled me to make the pre-predicate argument long enough to be able to clearly observe any associated intonational event. Furthermore, the target utterances had the same number of syllables in all of the stimuli. In this way, I ensured that any differences observed in the prosody were due to a difference in information structure.

Almost all of the people living in Santa María Tzejá where the experiment was carried out are non-literate in K'iche'. Therefore, it was not possible to use written material in the design of the experiment. Rather, the stimuli were recorded as conversations that the participants could listen to. In order not to bias the participants with native speaker

prosody, all of the target sentences were recorded by two non-native speakers of K'iche'. The questions, on the other hand, were recorded by a native speaker of K'iche'. Hence, all of the conversations were between a native and a non-native speaker of the language. To reduce the memory load, each conversation was accompanied by visual stimuli, e.g. pictures of women preparing beans and tortillas for the examples in (37) and (38). For each target sentence, the auditory and visual stimuli were the same across conditions. Figures 1 and 2 show the setup used for the contexts given above.

Maria

Figure 1: An example visual stimulus for focus

Manuela

Maria

Figure 2: An example visual stimulus for contrastive topic

4.2.2 The procedure

The participants were told that they were going to listen to conversations that consisted of question-answer pairs between a native speaker and two non-native speakers of K'iche'. They were also told that the non-native speakers were interested in hearing how native speakers would say the answers in the conversations. In Figures 1 and 2, clicking on the loudspeaker on the left played the conversation as a whole and clicking on the loudspeaker on the right played the same conversation without the target sentence. The task for each participant was first to listen to each conversation as a whole 1-2 times. Then the participant would listen to the same conversation one more time where the target sentence was removed and repeat the last sentence of the initial conversation as an answer to the question asked in the second conversation.

The participants were seated at a table in front of a laptop. Each participant wore head-mounted Sennheiser HMD280 headphones with microphone. The recordings were made with an Edirol R-09 recorder. 26 out of 192 utterances were excluded due to disfluency and the remaining 166 were included in the prosodic analysis.

4.3 Results

The research question I started out with was whether string-identical sentences with different meanings, namely focus and contrastive topic, also differ in their prosodic properties. The previous literature on contrastive topic and focus in K'iche' discussed whether constituents bearing these discourse functions are set off from the rest of the sentence by a pause. In order to see if this claim holds for the data I collected, each utterance was divided into two parts: (i) the pre-predicate part and (ii) the post-focal or post-topical part. I start with a discussion of the prosody of the pre-predicate expressions.

In all of the target utterances, the pre-predicate expression contained a rising pitch movement associated with the stressed syllable of the proper name. This finding is in line with the previous literature, in particular with Nielsen (2005) who claimed that K'iche' has stress-driven pitch accent where L+H* is the default pitch accent on content words.

In 50 (out of 166, ≈30%) utterances, the pre-predicate expression was followed by a pause. In the analysis, any physical pause between the pre-predicate expression and the rest of the sentences was taken into consideration. Such a pause after the pre-predicate expression was a proper pause and did not involve a stop closure because the following expression was always a verbal complex which began with a [ʃ] (*x* in the K'iche' orthography which stands for the completive aspect marker). The 50 pauses were distributed among the two conditions as follows: 29 focus, 21 contrastive topic. The mean duration of the pauses was 0.126s for focus and 0.115s for contrastive topic. A linear mixed effects model with speaker and item as random variables and condition as an independent variable did not yield any significant effect of condition. This finding shows that there is no clear indi-

cation that contrastive topics are distinguished from foci by a pause following them which goes against the claim made by Can Pixabaj & England (2011). As regards the prosodic boundaries following pre-predicate expressions, 77 of the focused expressions (out of 83, 92.7%) and 76 of the contrastively topicalized expressions (out of 83, 91.5%) were marked by a H% boundary tone. This finding shows that the boundary tone following the pre-predicate expressions is not affected by condition. It also goes against the claim that focus constituents do not form their own intonational phrases (Henderson 2012).

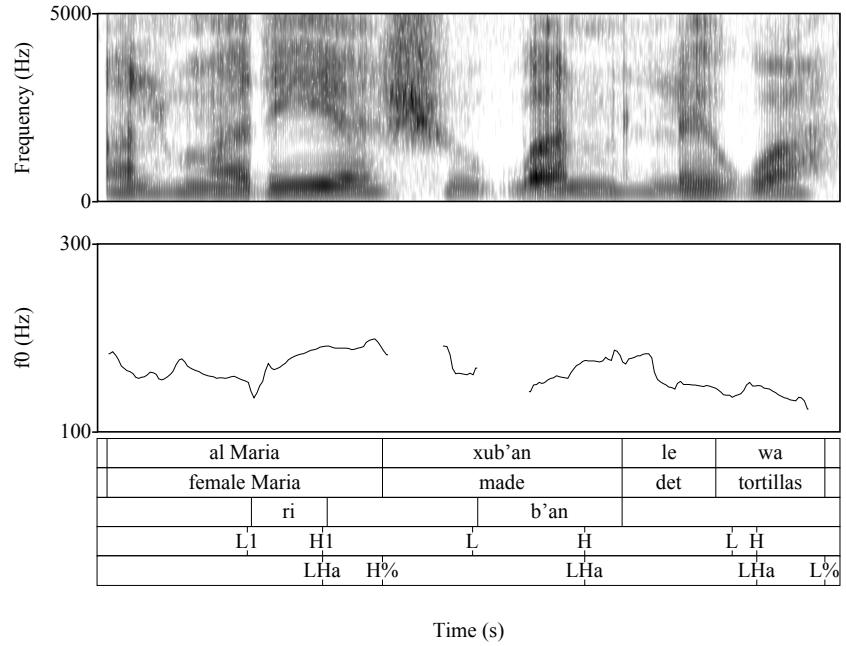

Figure 3: An example focus sentence

Following the previous work on the prosody of focus and contrastive topic that was discussed earlier, I have also looked at the following variables: (i) the duration of the stressed syllable in the pre-predicate expression, (ii) the alignment of the rising tone with respect to the onset of the stressed syllable, (iii) the duration of the rise of the F0 contour, (iv) the range of the rise, and (v) the slope of the rise. Figures 3 and 4 provide two example utterances illustrating how the analyses were carried out. In these figures, the first two tiers give the words and the glosses, respectively. The stressed syllables of the pre-predicate expression and the verb are marked in the third tier. The fourth tier provides information about the local minimum and maximum of the rises associated with each content word which are marked by the letters L and H, respectively. The last tier marks the rising pitch movement associated with each content word by LHa as well as the boundary tones, e.g. L% or H%.

Given these conventions about the annotation, Table 3 shows how the variables mentioned above are calculated. Here, I use t(x) to indicate the time corresponding to x

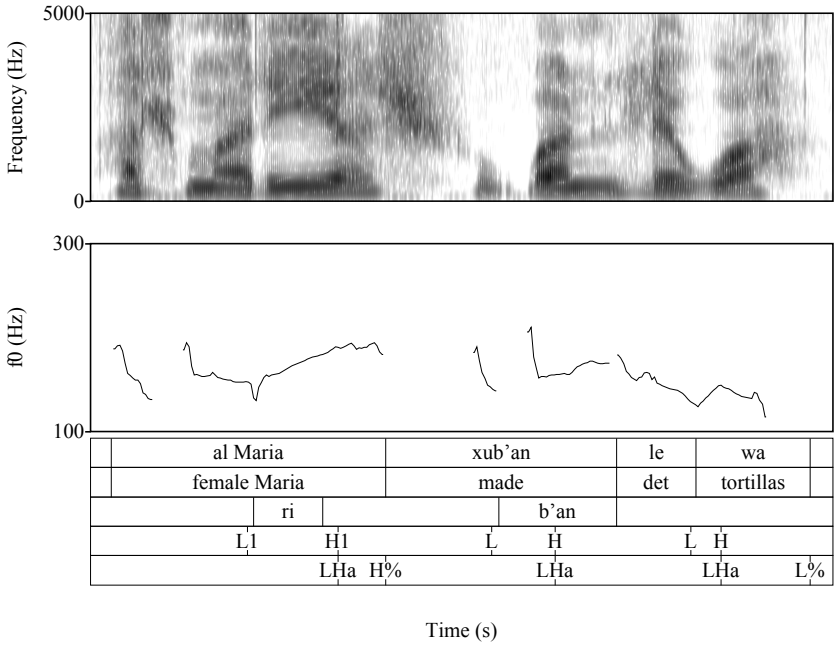

Figure 4: An example contrastive topic sentence

and $f_0(x)$ to indicate the F0 value corresponding to x:

Alignment of the L tone	t(L)-t(onset of the stressed syllable)
Alignment of the H tone	t(H)-t(onset of the stressed syllable)
Duration of the F0 rise	t(H)-t(L) (=D)
Range of the F0 rise	f_0(H)-f_0(L) (=R)
Slope of the F0 rise	R/D

Table 3: Calculation of the phonetic variables

Each dependent variable was fitted in a linear mixed effects model with speaker and item as random variables and condition as an independent variable. Among the measurements that were taken, there was a ∼6Hz difference in the range of the F0 rise across conditions and the linear models yielded a statistically significant result ($p < 0.05$) only for this variable. Figure 5 is a box plot that shows the distribution of the range of the rise on the pre-predicate expression across conditions. Given that this difference is very small, it may or may not be perceivable by K'iche' listeners. A perception study is needed to find out whether such a small difference is indeed perceivable.

I now turn to a discussion of the prosody of the post-focal or post-topical parts of the target sentences. A total of 104 utterances (64%) had a rising pitch movement on the verb. For these verbs, I carried out the same measurements as above. The remaining

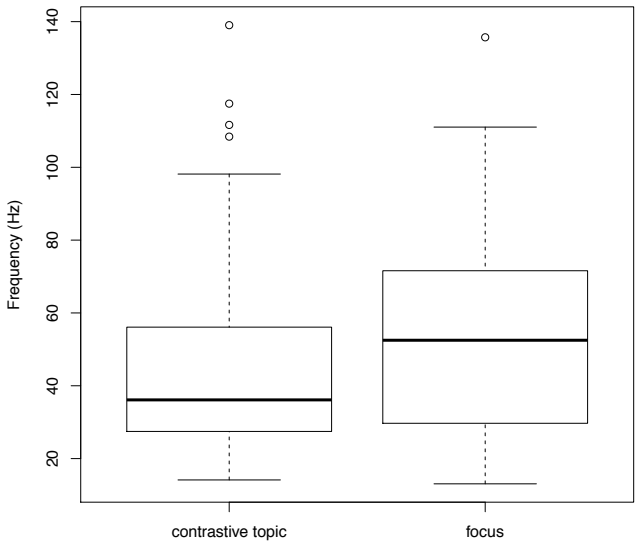

Figure 5: Range of the F0 rise by conditions

62 utterances did not have a rise on the verb either because (i) there was tonal truncation (n=54), or (ii) the F0 was flat on the verb (n=2), or (iii) the L or the H target could not be identified (n=6). For these cases, I only compared the H tone realized on the verb (if at all) across conditions. For each verb, I also looked at the duration of the stressed syllable.

As above, each dependent variable was fitted in a linear mixed effects model with speaker and item as random variables and condition as an independent variable. The linear mixed effects models did not yield significant effects for any of the variables.

5 Discussion

In the prosodic analyses of the data, each target sentence was divided into two parts, namely, a pre-predicate part and a post-focal or post-topical part, to be able to clearly see the predictions of the previous literature. Contrary to what Can Pixabaj & England (2011) claim, there is no clear indication that focus and contrastive topic are distinguished by a pause between the pre-predicate expression and the post-focal/post-topical expression. On the other hand, the data showed that there is a rising pitch movement associated with the focused or contrastively topicalized expression and a linear mixed effects model yielded a significant effect of condition on the range of this rise. However, the difference across conditions was small ($\sim$6 Hz) and requires a perception study to determine if such a difference matters for listeners. If the small difference in the F0 range that turned out to be significant in this study is actually not perceivable, then context may be the only source of

the intended interpretation.

In general, one can assume that pragmatic meanings are reflected in prosody because they are represented in the speaker's cognitive model and figure into speech planning. However, at a theoretical level, it is also possible that such an effect of pragmatic meanings on prosody may not always exist. Indeed, there are a series of studies on Yukatek Maya, e.g. Kügler *et al.* (2007); Kügler & Skopeteas (2006), which claim that there is no interaction between topic/focus and pitch manipulations. More generally, recent work suggests that there are languages where no prosodic reflexes of information structure are observed, e.g. Northern Sotho (Zerbian 2006), Hausa (Hartmann & Zimmermann 2007), Wolof (Rialland & Robert 2001) and Thompson River Salish (Koch 2008). A commonality across these languages is the use of word order changes and/or morphology to indicate the changes in information structure. The K'iche' data show that something similar might be going on in K'iche', especially if the significant difference in the range of the F0 rise is not perceivable by listeners.

6 Conclusion

This paper presented an experimental study on K'iche' designed to identify whether string-identical sentences with either a focus or a contrastive topic interpretation differ in their prosodic properties. The experiment involved obtaining naturally occurring data from native speakers of K'iche' by having them repeat target sentences they heard in conversations. The acoustic analyses of several variables yielded a significant effect of condition only in the range of the F0 rise associated with focused and contrastively topicalized expressions. However, the difference across conditions is only $\sim$6 Hz which may not be perceivable by listeners. Contrary to previous studies, the data did not support the claim that existence of a pause following the pre-predicate expressions distinguishes contrastive topics from foci.

The future research for this project can proceed in two directions. One is to improve the current experiment by using nominals in the pre-predicate position to prevent any interference of Spanish. The new stimuli should also include non-contrastively topicalized expressions in the pre-predicate position to determine the prosodic properties of such expressions and to compare them with the other two. Lastly, the experiment should be run on more participants in order to obtain a more representative sample. The second direction is to design a perception experiment where the data from the production experiment are used as stimuli. The perception experiment can be designed so that a given target utterance, say one where the pre-predicate expression is focused, can occur both in focus and contrastive topic contexts and the listeners can be asked to judge the acceptability of such utterances. If participants consistently accept a given target utterance in either context, then this shows that they do not make a prosodic distinction between focus and contrastive topic. Results from such an experiment would prove to be useful in interpreting the results of a production experiment especially if statistically significant but phonetically small differences are found.

References

AISSEN, JUDITH. 1992. Topic and focus in Mayan. *Language* 68.43–80.

——. 1999. Agent focus and inverse in Tzotzil. *Language* 75.451–485.

——. 2011. On the syntax of agent focus in K'ichee'. In *Proceedings of FAMLi I*, ed. by Kirill Shklovsky, Pedro Mateo Pedro, & Jessica Coon. MIT Working Papers in Linguistics.

BRAUN, BETTINA. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49.451–493.

BÜRING, DANIEL. 2003. On d-trees, beans, and b-accents. *Linguistics and Philosophy* 5.511–545.

——. 2011. Focus. In *The Cambridge Encyclopedia of the Language Sciences*, ed. by Patrick Colm Hogan. Cambridge: Cambridge University Press.

CAN PIXABAJ, TELMA. 2010. La predicación secundaria en K'ichee': Una construcción restringada. In *La predicación secundaria en lenguas de mesoamérica*, ed. by Judith Aissen & Roberto Zavala. Mexico City: CIESAS.

——, & NORA ENGLAND. 2011. Nominal topic and focus in K'ichee'. In *Representing Language: Essays in Honor of Judith Aissen*, ed. by Rodrigo Guitérrez-Bravo, Line Mikkelsen, & Eric Potsdam, 15–30. Linguistics Research Center, UC Santa Cruz Department of Linguistics.

COOPER, WILLIAM E., STEPHEN J. EADY, & PAMELA R. MUELLER. 1985. Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America* 77.21–42.

DAYLEY, JON P. 1981. Voice and ergativity in Mayan languages. *Journal of Mayan Linguistics* 2.3–82.

DIXON, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

EADY, STEPHEN J., WILLIAM E. COOPER, GAYLE V. KLOUDA, PAMELA R. MUELLER, & DAN W. LOTTS. 1986. Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29.233–251.

ENGLAND, NORA. 1991. Changes in basic word order in Mayan languages. *International Journal of American Linguistics* 57.446–486.

FACE, TIMOTHY. 2001. Focus and early peak alignment in Spanish intonation. *Probus* 13.223–246.

——. 2002. Local intonational marking of Spanish contrastive focus. *Probus* 14.71–92.

FÉRY, CAROLINE. 2006. The prosody of topicalization. In *On information structure, meaning and form: Generalizations across languages*, ed. by Kerstin Schwabe & Susanne Winkler, 69–86. Amsterdam, John Benjamins.

——, & VIERI SAMEK-LADOVICI. 2006. Focus projection and prosodic prominence in nested foci. *Language* 82.131–150.

GRICE, MARTINE, MARIAPAOLA D'IMPERIO, MICHELINA SAVINO, & CINZIA AVESANI. 2005. Strategies for intonation labelling across varieties of Italian. In *Prosodic typology*, ed. by S.-A. Jun, 362–289. Oxford: Oxford University Press.

HARTMANN, KATHARINA, & MALTE ZIMMERMANN. 2007. In place – out of place: Focus in Hausa. In *On information structure, meaning and form: Generalizations across languages*, ed. by Kerstin Schwabe & Susanne Winkler, 365–403. Amsterdam: John Benjamins.

HENDERSON, ROBERT. 2012. Morphological alternations at the intonational phrase edge. *Natural Language and Linguistic Theory* 30.741–787.

JACKENDOFF, RAY S. 1972. *Semantic interpretation in Generative Grammar*. Cambridge, MA: The MIT Press.

JUN, SUN-AH. 2005. Prosodic typology. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. by S.-A. Jun. Oxford: Oxford University Press.

KADMON, NIRIT. 2001. *Formal pragmatics*. Malden, MA: Blackwell.

KOCH, KARSTEN. 2008. *Intonation and focus in Nte?kepmxcin (Thompson River Salish)*. University of British Columbia dissertation.

KÜGLER, FRANK, & STAVROS SKOPETEAS. 2006. Interaction of lexical tone and information structure in Yucatec Maya. In *International Symposium on Tonal Aspects of Language*, 83–88, La Rochelle.

——, STAVROS SKOPETEAS, & ELISABETH VERHOEVEN. 2007. Encoding information structure in Yucatec Maya: on the interplay of prosody and syntax. *Interdisciplinary Studies on Information Structure* 8.187–208.

LARSEN, THOMAS W. 1987. The syntactic status of ergativity in Quiché. *Lingua* 71.33–59.

—— 1988. *Manifestations of ergativity in Quiché grammar*. University of California, Berkeley dissertation.

LÓPEZ IXCOY, CANDELARIA DOMINGA. 1997. *Ri ukemiik ri K'ichee' chii': Gramática K'ichee'*. Guatemala City: Cholsamaj.

MONDLOCH, JAMES L. 1978. *Basic Quiché grammar*. Institute for Mesoamerican Studies, State University of New York at Albany.

—— 1981. *Voice in Quiché-Maya*. State University of New York at Albany dissertation.

NIELSEN, KUNIKO. 2005. K'iche' intonation. *UCLA Working Papers in Phonetics* 104.45–60.

NORMAN, WILLIAM. 1977. Topic and focus in Mayan. In *Presentation at the Mayan Workshop II*, San Cristóbal de las casas, Chiapas, México.

PIERREHUMBERT, JANET. 1980. *The phonology and phonetics of English intonation*. MIT dissertation.

PYE, CLIFTON. 2001. The acquisition of finiteness in K'iche' Maya. In *Proceedings of the 25th Annual Boston University Conference on Language Development*, 645–656.

——, & PEDRO QUIXTAN POZ. 1988. Precocious passives (and antipassives) in Quiché Mayan. *Papers and reports on child language development* 27.71–80.

RIALLAND, ANNIE, & STÉPHANE ROBERT. 2001. The intonational system of Wolof. *Linguistics* 39.893–939.

RICHARDS, MICHAEL. 2003. *Atlas Lingüistico de Guatemala*. Guatemala City: Universidad de Rafael Landívar.

ROBERTS, CRAIGE. 1996. Information structure: Towards an integrated formal theory of pragmatics. In *OSUWPL*, ed. by Jae Hak Yoon & Andreas Kathol, volume 49, 91–136. The Ohio State University, Department of Linguistics. Reprinted with a new Afterword in *Semantics and Pragmatics* volume 5, 2012.

——. 1998. Focus, the flow of information, and Universal Grammar. In *The Limits of Syntax*, ed. by Peter Culicover & Louise McNally, 109–160. New York: Academic Press.

——. 2012. Topics. In *Semantics: An International Handbook of Natural Language Meaning*, ed. by Claudia Maienborn, Klaus von Heusinger, & Paul Portner, volume 33.2, 1908–1934. Mouton de Gruyter.

ROMERO, SERGIO. 2012. A Maya version of Jespersen's Cycle: The diachronic evolution of negative markers in K'iche' Maya. *International Journal of American Linguistics* 78.77–96.

ROOTH, MATS. 1992. A theory of focus interpretation. *Natural Language Semantics* 1.75–116.

——. 1996. Focus. In *The Handbook of Contemporary Semantic Theory*, ed. by Shalom Lappin. London: Blackwell.

STIEBELS, BARBARA. 2006. Agent focus in Mayan languages. *Natural Language and Linguistic Theory* 24.501–570.

TRECHSEL, FRANK R. 1993. Quiché focus constructions. *Lingua* 91.33–78.

VENDITTI, JENNIFER J., KIKUO MAEKAWA, & MARY BECKMAN. 2008. Prominence marking in the Japanese intonation system. In *Handbook of Japanese Linguistics*, ed. by Shigeru Miyagawa & Mamuro Saito, 456–512. Oxford: Oxford University Press.

WANG, BEI, & XI YU. 2011. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics* 39.595–611.

ZERBIAN, SABINE. 2006. *Expression of information structure in the Bantu language Northern Sotho*. Humboldt University dissertation.

The Ohio State University

Working Papers in Linguistics

No. 60

Edited by
Mary E. Beckman
Marivic Lesho
Judith Tonhauser
Tsz-Him Tsui

The Ohio State University
Department of Linguistics

222 Oxley Hall
1712 Neil Avenue
Columbus, Ohio 43210-1298 USA

Spring 2013

INTRODUCTION

This volume of the Ohio State Working Papers in Linguistics is a *varia* issue. It reflects a diversity of interests, containing papers that represent a good cross-section of the multiple sub-disciplines in the field of linguistics that are a focus of research in the department. Several authors even combine multiple sub-fields within the same paper. Rather than impose unwarranted interdisciplinary boundaries by attempting to group the papers by research area, we have instead ordered them simply by the alphabetical order of the (first) author's surname.

This volume is a landmark volume in two ways. First, it is volume 60, a special number by being five complete cycles of the Chinese zodiac. Second, it appears on the 50th anniversary of the first two degrees in linguistics to be awarded in the newly created Division of Linguistics at the Ohio State University, specifically the BA degree to Sandra A. Thompson and to Ray Saunders.

In June 2015, the department will celebrate the 50th anniversary of the establishment of doctoral program in linguistics at the Ohio State University. We are currently planning to publish, as volume 61 of *OSUWPL*, a special anniversary issue that will be a collection of papers by graduates of this program. This introduction also serves as a call for submissions to this volume. If you earned a Ph.D. from the Ohio State Department of Linguistics and would like to contribute a paper to the volume, please let us know by sending an e-mail to the editorial committee at:

wpl@ling.osu.edu

For subscription information and back issues, please visit us on the web at:

http://linguistics.osu.edu/research/publications/workingpapers

Ohio State University
Working Papers in Linguistics

No. 60

Table of Contents

MULTIPLE ANTECEDENT AGREEMENT AS SEMANTIC OR SYNTACTIC AGREEMENT[1]

Cynthia A. Johnson
Ohio State University

Abstract

In this paper, I challenge the argument put forth by Corbett (1991) that, within multiple antecedent agreement, the two possible agreement strategies, Resolution and Partial Agreement, can be viewed as semantic and syntactic agreement, respectively. Resolution, while semantically motivated and involving input from all of the agreement controllers, is not the same as semantic agreement in single-antecedent contexts. Partial Agreement, which relies on the morphological features of only one of the antecedents, still requires reference to the semantic features of both antecedents, as this strategy is more likely when the controllers are inanimate. Instead, I propose that the distribution of the two strategies – which nonetheless reflects the Agreement Hierarchy (Corbett 1979) and the Predicate Hierarchy (Comrie 1975) – is a product of the cognitive difficulty multiple antecedent agreement contexts pose for the speaker, such that the rules for this context are really part of broader principles within and across languages.

[1] An earlier version of this paper was presented at the 15th International Morphology Meeting in Vienna, Austria.

Cynthia A. Johnson. Multiple antecedent agreement as semantic or syntactic. *OSUWPL* Volume 60, Spring 2013, 1-9.

1. Introduction

Agreement with multiple antecedents provides surprising and interesting data for theories of syntax and agreement (Corbett 1991:261). In instances of multiple antecedent agreement, it is not immediately obvious from the features of the controllers alone what the features of the target should be. Unlike agreement with a single antecedent, where the controller and the target almost always share the same agreement features, agreement with multiple antecedents produces what Corbett (1991) refers to as an agreement mismatch between the controllers and the target(s).

Compare the following two examples from Latin. In example (1), agreement with a single antecedent, the controller is *Scipio*, a masculine singular noun. As expected, the targets share the same features as the controller: the verb *sit* agrees for the correct number, singular, and the adjective *clarus* agrees for both gender and number, masculine singular.

(1) Single Antecedent Agreement
sit Scipio clarus
be-3.SG Scipio.M.SG illustrious.M.SG
'Let Scipio be illustrious.' (Cicero, *Cat*. i*v.21*, from A&G)

On the other hand, in example (2), there are two antecedents (in the form of a conjoined noun phrase, *labor voluptasque* 'labor and desire') which control agreement on the target verb/past participle *esse iunctum*[2] 'be-bound.INF'. Since the target cannot be masculine (following *labor*) and feminine (following *voluptas*) at the same time, we have the precise context for an agreement mismatch: the features of the target will never match both controllers at the same time, since only one set of feature values can be expressed by a target. What we find instead is a small range of possible feature combinations for the target, modeled on either just one of the antecedents or according to a set of semantically-based rules.

(2) Context for Multiple Antecedent Agreement
Labor voluptasque ESSE IUNCTUM
labor.M.SG delight.F.SG.-and be bound
'Labor and delight are bound'

In Latin, and across languages more broadly, these possible feature combinations result from two main strategies for approaching multiple antecedent agreement: Resolution (3a) and Partial Agreement (3b) (terminology from Corbett 1991 and Wechsler & Zlatić 2003). Resolution appears to be more semantically motivated: the target's features are more or less "computed" by adding up the features of the controllers. As a result, the target is always plural (reflecting the semantic transparency of number agreement in this context) and the gender is determined according to language-specific

[2] This is not the usual citation form for Latin verbs (the first person singular present indicative is primarily used), but it represents a "neutral" form of the past passive verb: the auxiliary is in the infinitive (so as not to express gender) and the participle is in the neuter singular, the common citation form for adjectives.

rules, which often reference semantic features, e.g. animacy. Partial Agreement, on the other hand, does not display the same semantic motivation: rather than involving equal contribution of both of the controllers, the target's features match those of only one of the controllers. In Latin, the controller that serves as the basis for agreement is usually the one closest to the target (and so Partial Agreement in Latin is referred to as "Nearest Antecedent Agreement").

(3) Agreement Strategies

a. *Resolution*
formosi sunt verris et scrofa
handsome.M.PL are boar.M.SG and sow.F.SG
'The boar and the sow are handsome.' (Varro, *RR II.4.4*)

b. *Nearest Antecedent Agreement*
ut maxime **amicum** **cytisum** et medica
while very beneficial.N.SG snail-clover.N.SG and alfalfa.F.SG
'while snail-clover and alfalfa [are] very beneficial' (Varro, *RR II.2.19*)

One of the primary goals in investigating multiple antecedent agreement is to model the distribution of the strategies: within and across languages, in what contexts do we find Resolution and Partial Agreement? Corbett (1991) has proposed that the distribution of these strategies conforms to the Agreement Hierarchy (4a, Corbett 1979) and the Predicate Hierarchy (4b, Comrie 1975), if we view Resolution as semantic agreement and Partial Agreement as syntactic agreement.

(4) Typological Generalizations

a. *The Agreement Hierarchy*
attributive | predicate | relative pronoun | personal pronoun
⟵ syntactic agreement — semantic agreement ⟶
Nearest Antecedent Agreement — Resolution

b. *The Predicate Hierarchy*
verb | participle | adjective | noun
⟵ syntactic agreement — semantic agreement ⟶
Nearest Antecedent Agreement — Resolution

These hierarchies provide a basic typology of agreement: the more noun-like the target, the more likely we are to find agreement with the semantic features of the controller(s); the more verb-like the target, the more likely we are to find agreement with the syntactic (i.e. grammatical) features of the target.

Corbett's proposal, while supported by cross-linguistic research and single antecedent agreement structures, raises an important theoretical question: is Resolution really semantic agreement and is Partial Agreement really syntactic agreement? Is there independent evidence for viewing the strategies in this way, and should we expect multiple antecedent agreement to operate in the same way as the straightforward single antecedent agreement in (1)? In what follows, I provide a critique of Corbett's proposal

using data from Latin and suggest an alternative solution: one that takes into consideration linguistic performance. In sections 2 and 3, I discuss Resolution and semantic agreement, and Partial Agreement and syntactic agreement. In section 4, I present my performance-based view of agreement, which accounts for the distribution of strategies with reference to broader principles and rules within and across languages.

2. Resolution and semantic agreement

In Latin, there are two patterns of Resolution: in some instances, masculine is the resolved gender (5, repeated from (3a) above), but in other instances, neuter is the resolved gender (6).

(5) formosi sunt verris et scrofa
handsome.**M.PL** are boar.M.SG and sow.F.SG
'The boar and the sow are handsome.' (Varro, *RR II.4.4*)

(6) labor voluptasque … sunt iuncta
labor.M.SG delight.F.SG.-and are bound.**N.PL**
'labor and delight are bound...' (Livy. *AVC*, from A&G)

In both examples, there is a masculine singular antecedent and a feminine singular antecedent, i.e. the same grammatical features are present for both examples. The Resolution rules must therefore refer to a feature other than the grammatical gender of the antecedents to determine the resolved gender for each sentence.

The relevant feature is animacy, as discussed by the grammar handbooks (e.g. Allen & Greenough 1888) and previous literature (e.g. Corbett 1991), and supported by the data in my own corpus study (Johnson 2011). Animate antecedents use masculine as the resolved gender, while inanimate antecedents use neuter as the resolved gender. Given this connection between animacy and resolved gender, there is a clear semantic basis to these Resolution rules: at the very least, there is a connection between neuter grammatical gender and lacking biological gender and, perhaps to extend this connection, between masculine grammatical gender and having biological gender.

We have established that Resolution is semantically motivated, involving contribution of the semantic features of both antecedents, but is this actually semantic agreement? This would imply that the resolved gender follows naturally from the "adding up" of the controllers' gender features. Consider example (7) of semantic agreement in a **single** antecedent context.

(7) pars certare parati
part.F.SG to contend ready.M.PL
'a part [group of men] ready to contend'[3] (Vergil, *Aen*. v. 108)

[3]The larger context (7): laeto complerant litora coetu / uisuri Aeneadas, pars et certare parati.
'They filled the shores with a happy crowd / [some] to see the men of Aeneas, and a part ready to contend [in the games].' (Aeneid, v. 107-8)

In this example, the target *parati* is masculine plural, even though the controller, *pars*, is grammatically feminine singular. This is because *pars* refers to a group of men out in the world: the grammatical features of *parati* are modeled on the plurality and male-ness of the group of men that is the referent of *pars*.

The question we should consider is whether Resolution works in the same way: do the features of the target follow naturally from the "adding up" of the semantic features of the controllers? There are at least two problems here: first, there is necessarily a stipulated component to these rules. Resolution rules vary across languages: in Old Icelandic, for example, we find that all instances of Resolution are to the neuter gender, regardless of the semantic properties of the antecedents (Corbett 1991:80-3). If the gender of the group follows naturally from the semantics of the controllers, we would not expect to see this kind of variation. Additionally, the process of "adding up" genders to produce masculine or neuter gender is not semantically transparent. How does masculine and feminine combine to "equal" masculine or neuter gender? There is no transparent connection between the semantic genders of the controllers and the resulting resolved gender[4].

While Resolution is unquestionably semantically *motivated*, it is not the same kind of semantic agreement found in single antecedent agreement. Our explanation for the distribution of the two strategies should therefore reflect this fundamental difference between single antecedent agreement and multiple antecedent agreement.

3. Partial Agreement and syntactic agreement

The other multiple antecedent agreement strategy in Latin, Nearest Antecedent Agreement, occurs when the target shares the same feature values with only the closest controller, regardless of if the target follows (8a) or precedes (8b, repeated from (3b) above) the controllers. In cases where there is more than one agreement target, each target agrees with its closest controller (8c).

(8) a. Ibi Orgetorigis filia atque **unus** e filiis
There of-Orgetorix daughter.F.SG and one.**M.SG** from sons

captus est
was-captured.**M.SG**
'There the daughter and one of the sons of Orgetorix were captured.'
(Caesar, *BG 1.26*)

b. ut maxime **amicum** **cytisum** et medica
while very beneficial.**N.SG** snail-clover.**N.SG** and alfalfa.F.SG
'while snail-clover and alfalfa [are] very beneficial' (Varro, *RR II.2.19*)

[4] It was suggested (Corbett, p.c.) that the key to the semantics might lie in the Latin word(s) for 'group'; however, all three genders are represented by the various Latin 'group' words: *coetus* 'assembly' (m.), *classis* 'group, division' (f.), *decuria* 'gang, class' (f.), *conjectus* 'throwing together, i.e. collection' (m.), *collectio* 'collection' (f.), *corpus* 'body' (n.).

c. non **eadem** **alacritate** ac **studio** **quo**
not same.**F.SG** ardor.**F.SG** and zeal.**N.SG** which.**N.SG**
'[did not employ] the same ardor and zeal which [they had used to employ in land combat]' (Caesar BG. 4.24)

Again, we can ask a similar question: is Nearest Antecedent Agreement syntactic agreement? Syntactic agreement is defined as agreement consistent with the morphological features of the controller(s), without reference to the semantic features (Corbett 2006:156). In Nearest Antecedent Agreement, the target's features are only consistent with the morphological features of *one* of the controllers – and it is the controller that is nearer to the target. This local and linear dimension to Nearest Antecedent Agreement should not be ignored; as discussed below, Nearest Antecedent Agreement resembles a typical "ungrammatical" outcome of difficult long distance dependencies, i.e. attraction errors.

Additionally, while syntactic agreement implies no reference to the semantic features of the controllers, Corbett (1991) has stated that Resolution is more likely when the controllers are animate, i.e. there is some reference to a semantic feature when "choosing" which strategy to use.

Viewing Nearest Antecedent Agreement as syntactic agreement works only if this strategy is completely divorced from the semantics. In a broad sense, this is appropriate: there is no semantic rule that conditions the form of the target; it is only the proximity of one of the controllers that determines the features. However, we still need to explain why such a strategy does not involve input from all of the controllers (there could just as easily be a syntactic rule that computes the gender of the target from the morphological features of the controllers). Likewise, the semantic features of the controllers still influence the choice of strategy, even though the actual agreement process of Nearest Antecedent Agreement does not make reference to any semantic feature. This aspect of the distribution also requires explanation, as it is not explained by the hierarchies above.

4. Performance-based agreement: Gender assignment and Avoidance

If, on the basis of these facts, Resolution is not quite semantic agreement and Partial Agreement is not quite syntactic agreement, we need to explain why the hierarchies in (4) are still observed in Latin. In fact, even if Resolution **is** semantic agreement and Partial Agreement **is** syntactic agreement, such patterns still require explanation: the hierarchies in (4) are only typological tools that model common cross-linguistic patterns; by themselves, they offer no explanation as to why such patterns frequently occur. The solution I propose is one that takes into consideration linguistic performance, as evidenced in particular by the existence and acceptability of a strategy like Nearest Antecedent Agreement.

Unlike Resolution, Nearest Antecedent Agreement relies on linear and local relationships between the controllers and the target(s). As mentioned above, this strategy

– at least superficially – resembles what are typically referred to as attraction errors in other languages, e.g. the examples in (9) below.

(9) a. *Number Attraction*
The time for fun and games **are** over. (Bock & Miller 1991)

b. *Gender Attraction*
Stanze che sono anni e anni che sono
rooms.F.PL that be.3.PL years.M.PL and years M.PL that be.3.PL.

chiusi
closed. M.PL
'Rooms that have been closed for years and years' (Vigliocco & Franck 1999)

Along the same lines, Corbett (2006:170) has argued that, with respect to this agreement strategy, we should "perhaps be looking to psychologists, who have demonstrated the importance of first and last positions in lists in other domains."

There is also an inherent difficulty present in multiple antecedent agreement contexts. First, as discussed earlier, both strategies produce an agreement mismatch: the target cannot share the same features as both of the controllers. This makes the task of selecting agreement features more complicated than in single antecedent agreement contexts. Second, the gender system in Latin may be another source of difficulty for speakers. For animate nouns, the grammatical gender of the noun overlaps with the biological gender of the referent. This creates a gender system in Latin that in some instances references the natural sex of the referent, but in other instances, e.g. for inanimate nouns, it does not–it is purely grammatical, and thus has no relationship with the actual semantic properties of the referent.

Finally, multiple antecedent agreement is relatively rare: in my 300,000-word corpus study (Johnson 2011), there were only 47 unambiguous tokens, which means that speakers encounter this context far less than they do single antecedent agreement contexts. All of these facts about multiple antecedent agreement and Nearest Antecedent Agreement indicate – at least indirectly – that such a construction causes the speaker cognitive difficulty. The resulting strategies are a product of this difficulty, such that the strategies are a result of agreement done "on the fly," according to more general rules within and across languages.

4.1 Resolution as gender assignment

Resolution is simply gender assignment. Within Latin, both semantic and formal criteria are relevant for gender assignment: both the semantic features of the noun and the form of the ending are used for gender assignment, e.g. for borrowed words. In multiple antecedent agreement, the targets are all native words, so only the semantic criteria are used. In particular, it is the animacy value of the nouns that determines the assigned

gender. Wechsler & Zlatić (2003:182-3) have formalized this notion by proposing that coordinate noun phrases do not have an inherent lexical gender feature and so must be assigned a semantic gender based on a language-specific rule. In Latin, this rule – which operates not just in multiple antecedent contexts but elsewhere in the language – is one that simply correlates masculine grammatical gender with animacy and neuter grammatical gender with inanimacy. Since this rule applies in other feature assignment contexts, we are able to explain Resolution with reference to a broader rule within Latin.

4.2 Nearest Antecedent Agreement as Avoidance

Alternatively, rather than dealing with the complex problem of "adding up" genders via Resolution/gender assignment, the speaker can choose to avoid this problem altogether by simply agreeing with the closest antecedent. Nearest Antecedent Agreement is thus part of a larger category of what Hock (2007a) terms "Avoidance" strategies, whereby speakers employ a strategy that does not require them to produce a resolved form (where they must address the difficulty posed by the resulting agreement mismatch, the lack of semantic transparency in the gender marking, and the infrequency of the agreement context). Other Avoidance strategies found across languages include First Antecedent Agreement (e.g. in Slovene, Corbett 1991:266), restructuring the sentence completely (e.g. in Polish, Rothstein 1993), and gender neutralization (e.g. in German, Hock 2007b).

How does this performance-based view of multiple antecedent agreement fit with the hierarchies in (4) above? Rather than labeling Resolution as "semantic agreement" and Nearest Antecedent Agreement as "syntactic agreement," I account for the distribution of Resolution vs. Nearest Antecedent Agreement as one that is the product of the relative difficulty of different multiple antecedent agreement contexts. Resolution occurs when the semantics of the antecedents are more concrete (when the controllers are animate) and/or more relevant (when the target is more noun-like, i.e. when we must conceive of the coordinate noun phrase as a group).

Nearest Antecedent Agreement, on the other hand, is a product of the cognitive difficulty such contexts create, especially when the semantic features are less transparent (when the controllers are inanimate) and/or less relevant (when the target is more verb-like). The hierarchies are therefore explained not with reference to semantic or syntactic agreement – a problematic notion given the facts above – but according to the difficulty posed by multiple antecedent agreement more generally.

5. Conclusion

The Agreement Hierarchy and the Predicate Hierarchy are useful typological tools: in single antecedent agreement contexts, they accurately describe how likely a speaker is to agree with the semantic or morphological features of the controller. In multiple antecedent agreement, the same patterns are observed, provided we put Resolution on the semantic agreement end of the hierarchy and Partial Agreement on the syntactic agreement end. However, this conceptualization of the agreement strategies is met with significant theoretical problems: Resolution is not quite semantic agreement as defined in

single antecedent agreement contexts, and Partial Agreement still involves some reference to the semantic features of the controllers.

In order to account for the observed distribution of strategies, I instead propose that the patterns are a result of the overall cognitive difficulty associated with such an infrequent structure – a structure that is further complicated by the nature of gender marking in Latin and the agreement mismatch that necessarily results from multiple antecedent agreement. In this way, the strategies can be explained by broader principles within and across languages: Resolution as gender assignment, and Partial Agreement as a kind of Avoidance strategy.

References

Allen, Joseph Henry, and James Bradstreet Greenough. 1888. *New Latin grammar for schools and colleges.* Boston: Ginn and Company.

Bock, Kathryn, and Carol A. Miller. 1991. Broken agreement. *Cognitive Psychology* 23:45-93.

Comrie, Bernard. 1975. Polite plurals and predicate agreement. *Language* 51(2):406–18.

Corbett, Greville. 2006. *Agreement.* Cambridge, UK: Cambridge University Press.

Corbett, Greville. 1979. The agreement hierarchy. *Journal of Linguistics* 15(2):203-224.

Corbett, Greville. 1991. *Gender*. Cambridge, UK: Cambridge University Press.

Hock, Hans Henrich. 2007a. Agreeing to disagree: Agreement with non-agreeing antecedents, with focus on Sanskrit and Latin. East Coast Indo-European Conference, Yale University, June 2007.

Hock, Hans Henrich. 2007b. Early Germanic agreement with mixed-gender antecedents with focus on the history of German. Paper presented at the UCLA Indo-European Conference, 2-3 November 2007.

Johnson, Cynthia A. 2011. Multiple antecedent agreement in Latin. Paper submitted as the first qualifying paper for completion of doctoral program in Linguistics. Ohio State University Department of Linguistics.

Rothstein, Robert A. 1993. Polish. In Comrie, Bernard & Greville G. Corbett, eds. *The Slavonic languages*. London/New York: Routledge, 686-758.

Viggliocco, Gabriella, and Julie Franck. 1999. When sex and syntax go hand in hand: gender agreement in language production. *Journal of Memory and Language* 40:455-78.

Wechsler, Stephen, and Larisa Zlatić. 2003. *The many faces of agreement*. Stanford: CSLI Publications.

ON PHONICALLY BASED ANALOGY*

Brian D. Joseph
The Ohio State University

Abstract

In this paper I examine the role sound alone can play as the basis for analogical connections among forms, as opposed to more conventionally discussed factors such as paradigmatic structure, grammatical category, or meaning. Examples are presented here, mainly from English, that show sound effects in analogy at various levels of linguistic analysis, including phonetics, morphology, syntax, semantics, and the lexicon.

1 Introduction

Analogy, understood here in a broad sense to refer to any change in a given form due to the influence of another form, has a venerable history of study within linguistics, dating back to the Greek and Roman grammarians and their interest in the relationship between analogy and the origin of words and the origin of language itself. It is not surprising, therefore, that various textbooks on historical linguistics, perhaps most notably Anttila 1972/1989, have made clear the prominent role that analogy plays in the understanding of

* The material in this contribution is drawn from a presentation I have made in numerous venues since 2001 under various titles — too many to list — but beginning when I was a fellow at the Research Centre for Linguistic Typology at La Trobe University in July and August of 2001, at the kind invitation of R. M. W. Dixon and Alexandra Aikhenvald. I gratefully acknowledge the invaluable support of my residence there to this work, and thank the various audiences over the past few years who have contributed important insights to my thinking on the examples discussed herein.

Brian D. Joseph. On phonically conditioned analogy. *OSUWPL* Volume 60, Spring 2013, 11-20.

language change. Anttila's work, elaborated upon in Anttila 1977, established (perhaps, re-established) the semiotic underpinnings of analogical change.[1]

Still, even with so much attention to the topic, questions remain about analogy. One such question, given that analogy depends on a connection being made between two forms (the influencer and the influencee, so to speak), is just what sorts of connections can serve as the basis for analogical pressures and ultimately for analogical re-formations.

In this brief piece, I present a number of examples I have collected over the years that address this key question by demonstrating that one type of linkage between forms that must be recognized is a purely phonic one, based on sound alone. This is so even though sound is not generally thought of as a basis for analogical connections; most discussions of analogy in historical linguistics textbooks focus only on grammatical connections between forms, e.g. forms that are in the same paradigm (traditional "leveling" or "internal analogy") or forms that are members of the same grammatical category ("form class analogy" or "external analogy").

The general neglect[2] of a phonic basis for analogy is perhaps somewhat surprising, given that a phonic basis can be found in other aspects of language use. For instance, sound is critical in many types of language play, among them counting rhymes, such as *eeny, meeny, miny, mo* with its assonance and alliteration. Moreover, sound plays an important role, beyond simple rhyming patterns, in various sorts of literary expression; for instance, Miller 1982 has demonstrated complex phonic echoing within lines in Homeric epics, Watkins 1995 has shown the importance of phonic devices linked to thematic parallels throughout several ancient Indo-European poetic traditions, and Dawson 2005 draws attention to the effects of homoioteleuton, a phonically based poetic (and rhetorical) device, in the selection of certain dual and locative allomorphs in Vedic Sanskrit.[3] Further, even within recognized types of analogy, a phonic basis often is lurking. For instance, classic cases of 'contamination', which in one sense can be viewed as leveling within a 'semantic paradigm', can involve a phonic link. A relevant example is Late Latin *grevis*, which is generally believed to have developed from Classical Latin *gravis* "heavy" through 'contamination' with its semantic opposite *levis* "light", plus some influence likely from the semantically related (as a dimension adjective) *brevis* "short; brief"; however, even if the semantic links were important here — and I have no doubt that they were — there is a phonic link as well with *gravis/grevis*, *levis*, *brevis*, in

[1] Note also the excellent bibliography on analogy, Anttila & Brewer 1977, and various recent handbook-style treatments of analogy, especially Anttila 2003 and Hock 2003.

[2] There are exceptions; Vennemann 1972, for instance, with its discussion of 'phonetic analogy', clearly emphasizes that the notion of analogy must be extended to include connections made at the level of sound and not of grammar proper. Claims concerning the purely grammatical basis of analogy are to be found in work done within the framework of Optimality Theory, on 'correspondence theory', in that the typical basis for correspondence relations is grammatical outputs, forms being considered by the evaluation mechanism of the grammar.

[3] Relevant here too is what Hock and Joseph (1996: 293), drawing on the fine work of Samuels 1972, call 'phonesthematic attraction' to describe cases where sound symbolic elements attract other forms into taking on some aspect of their shape (as with early Modern English *sacke* "sink, droop" turning into *sag* through the influence of other words in [-æg] with meanings pertaining to "slow, tiring, tedious action"); since sound symbols can potentially be considered morphemic in nature, the influence in such cases is not just phonic but involves some semantic basis as well.

that they all share the phoneme sequence -VOWEL–*vis* (of which the *–vi-* can be considered a shared stem-forming morpheme).

2 Case Studies in Phonically Based Analogy

The examples presented here range over changes in pronunciation (2.1-2.5),[4] changes in meaning (2.6-2.8), including an example from language contact/bilingualism, changes in lexicon and morphology (2.9-2.10), and changes in syntax (2.11-2.12). In many, perhaps most, of these cases, it is not possible to demonstrate conclusively that sound alone is responsible for the change (though 2.5 comes close), but the aggregate effect of so many examples in which sound seems to have been a relevant dimension to the analogical linkage, I would claim, is to show that a phonic basis for analogy is a distinct possibility that cannot simply be dismissed and thus must be taken into consideration whenever analogy is invoked.

All of the forms cited here are ones that I have heard over the past 30 or so years of collecting interesting examples of language change in action. Although I cannot give precise information about the speakers or the circumstances under which the form was uttered, I vouch for the accuracy of my noting of the forms and note that none is based on a unique instance, and some may even represent longer-standing variation that has been maintained.[5] In each case, I present the facts along with my interpretation of a phonic basis for the analogy, offered without an extensive justification at this stage, in hopes of sparking the necessary weighing of alternative interpretations. Also, where the examples provide the basis for some observations of a more general kind about the nature of analogy, some further comments are included.

2.1 Modern English <memento>

A common pronunciation for the word *memento* "a reminder of the past" in modern American English is [momento] with [o] in the first syllable instead of the 'correct', i.e. historically prior and otherwise expected (note the spelling, for instance) mid-vowel [e]. No similar change is observed in the word *pimento* nor, perhaps more importantly since it involves the same morpheme, in *memorial*, suggesting that the change in *memento* cannot be a regular sound change affecting [e] or [I] between labials, for instance. Presumably, the [o] is based on the word *moment*, which is strongly linked phonically with *memento* due to their sharing the onset of *mVm* and to their both having the sequence *–nt-* following later in the word. Admittedly, there is also a weak semantic link via the phrase *of great moment* and the adjective *momentous***,** both of which mean "memorable" to some extent. More interestingly, one effect of the phonic analogy that leads to [momento] is a severing — or at least weakening — of the once-phonetically compatible linkage between *memento* and *memorial* and other derivatives, or to put it in a different way, the

[4] Note that changes in pronunciation are not the same as sound change, as they may have a variety of causes, including nonphonetic ones, and they need not be regular; I take regularity and purely phonetic conditioning to be the hallmarks of sound change in the strict sense, what may be called 'Neogrammarian sound change' or 'sound change proper' (see Joseph 2008 and Anderson, Dawson, and Joseph 2010: 267 on this latter term).

[5] Such is the case with 2.6 (*flaunt* vs. *flout*), as Henning Andersen (personal communication, 5 October 2004) has brought to my attention (and cf. the OED's citation of *flaunt* in the sense of 'flout' from as early as 1923); so also with 2.3 (*nuclear*) and 2.11 (the *as far as* construction), and possibly others.

morphemic linkage with *memorial*, *remember*, etc. was not strong enough to counteract the effect of the phonic linkage with *moment*.

2.2 Modern English <consonantal>

The adjective associated with the noun *consonant* is *consonantal*, meaning "having to do with a consonant", and while it is generally pronounced, as would be expected from the spelling, [kansənæntəl], there are speakers, such as myself, who regularly say instead [kansənɛntəl]. The source of this innovative pronunciation is obscure, to be sure, but it is presumably based on near-rhyme *continental*; there is here some morphological link in that both *consonantal* and *continental* are denominal adjectives in *–al*, but the main connection between the two is sound-based, via shared onset, shared syllable-count, and shared syllable structure. Moreover, as with *grevis* discussed above, even a morphemic link gives a phonic link, here with respect to the final element *–al*.

2.3 Modern American English <nuclear>

One relevant case that has gotten a fair bit of play over the years in the popular press due to its being, it seems, the pronunciation of choice among American presidents, including Dwight D. Eisenhower, Jimmy Carter, and George W. Bush, is the adjective *nuclear* "having to do with a nucleus" pronounced as [nukyulər], for etymologically correct [nukliyər]. Here the influence seems to be the class of adjectives like *popular*, *particular*, *insular*, etc., with *nuclear* in essence 'assimilating' to, i.e. being attracted into, the class of adjectives in *–lar*. However, even if the end-point is a morphological type with a suffixal *–lar*, the starting point has to be the phonic form, with no strong morphemic basis. That is, even though *nucleus* has an *–l-* in it, its *–l-* has a different placement and morphemic status from that seen in *people/populace*, *particle*, etc., i.e. in the base words for *popular*, *particular*, etc.[6] The phonic form that gives a starting point for the attraction is *...lər* in both the attractor and attractee, discontinuous in the case of *nuclear* (thus *...l...ər*), and the end result is *...(yu)lər* in both.

2.4 Modern American English <extraterrestrial>

The adjective *extraterrestrial* "from outer space" is innovatively pronounced by some speakers as ending in [...stiyəl] as opposed to the etymologically correct ending [...striyəl]); the basis here seems to be attraction to, that is to say influence from, *celestial* 'heavenly', with the phonic link being the shared sounds [–ɛst…iyəl], though admittedly there is a semantic connection as well between these words.

2.5 Modern American English <academia>

One particularly intriguing case is the pronunciation of *academia* as [ækəd**ej**miə] (at least in American English) as opposed to the more usual [ækəd**ij**miə]. In talking about this case over the years, in classes or in presentations, I have been told that it is a pseudo-

[6] Dr. Tom Stewart (personal communication, Spring 2001) has told me that the noun *nucleus* can be heard as [nukjuləs], and I have personally verified that since. Though this could be the basis for the adjectival pronunciation discussed here, I am inclined to think — since there is no obvious (to me) basis for [nukjuləs] in and of itself — that the noun here is a back-formation derived from the innovative pronunciation of the adjective.

learnedism, affecting a Latin-like style of pronunciation or an Italian- or a Spanish-like one, but that ignores the basic point of why this word out of other possible words would have been affected, and why that particular affectation as opposed to other possible alterations occurred with this word. That is, there are other learnèd words that do not undergo a similar fate, such as *anemia*, for which there is no variant [ən**ej**miə],[7] or even *epidemiology*, with the same surrounding environment as *academia* (i.e., with *d* and *m* flanking the affected vowel), for which there is no [ɛpɪd**ej**mi…].[8] Nor can influence from a morphologically or semantically related word be responsible; in fact, one does hear on occasion [ækəd**ɛ**miə], based on the pronunciation of *academic*, but there is no obviously related word with [-ej-]. But when one looks to less obvious (but, I would argue, no less relevant) forms, a solution awaits; thus, I suggest that this innovative pronunciation of *academia* is based on the influence of *macadamia (nut)* where the basis for the connection is purely phonic in nature – the relevant phonic links are the large number of shared segments in the same order, in particular, [ækəd…miə]) and the shared rhythmic stress pattern. This influence seems to be felt even though there is no semantic connection whatsoever; phonic form alone seems to matter here.[9]

2.6 American English <flaunt>

The verb *flaunt*, canonically having the meaning "show off; display ostentatiously", can now be used as well quite commonly (though prescriptively 'incorrectly') in meaning of "show contempt for; scorn". This innovative meaning is exactly the meaning of *flout*, which, not coincidentally I would argue, is phonically similar to *flaunt* in that both share [fl…t]. Thus, with this innovative meaning, *flaunt* has assimilated in meaning to *flout*, where the link between the two, the basis for the analogical influence of *flout* over *flaunt*, is a shared aspect of their phonic shape.

2.7 American English <diffident>

Somewhat similar to *flaunt* is the situation with *diffident* 'shy, lacking in self-confidence", in that it is now used by some speakers in the meaning of "having no interest in or concern for". Presumably what has happened here is that *diffident* has been 'attracted' by the phonically similar *indifferent*, which has that very meaning; crucial here to the attraction is the fact that the two words share the syllable [...dɪf...]), which is

---

[7] My good friend and many-time collaborator Richard Janda and I independently came up with this idea about the source of the innovative pronunciation of *academia*, at some point in the mid-to-late 1980s, and we have each since used it in classes and in presentations. My including it here in print is with Rich's permission, and in fact, I must acknowledge his input through enlightening discussion we have had on this example, including the particular point about *anemia*; I have benefited greatly too from the many general discussions Rich and I have had over the years concerning not just *academia* but the whole overall line of reasoning adopted herein as well.

[8] Henning Andersen (personal communication, 5 October 2004) tells me that the word *schizophrenia*, widely pronounced as ending in […ijniə], can be heard also as ending in […ejniə], suggesting that there may indeed be a 'learnèd word' pronunciation coming to be associated with [ej] in certain items. I am inclined however to think of possible influence from semantically (and somewhat phonically) connected *mania* in this case, though the nature of this sort of variation in general is such that one cannot rule out any of the possible pressures.

[9] One is inevitably led to make a quip about academics being nuts, and indeed, I even own a T-shirt, a gift from a former student, Dr. Halyna Sydorenko of Toronto, that says "Academia Nut". Such a connection seems unlikely to have played a role here, however. (To see a photo of me in the shirt, go to http://osu.academia.edu/BrianDJoseph.)

stressed in each, as well as having the same end segments and, except for the prefix *in-*, the same rhythmic structure.

2.8 American Norwegian <brand>

The same effect as that seen in 2.6 and 2.7 can be observed in language contact, where the 'attraction' takes place across languages whose speakers are in contact and are bilingual. In particular, Haugen 1969 has noted what he calls 'homophonous extensions', exemplified by American Norwegian *brand*, which has the meaning "bran", as opposed to the meaning "fire" in Standard Norwegian, due, in his account, to the influence of American English *bran*. Haugen's use of the descriptor 'homophonous' signals his recognition of the relevance of the phonic link between the attractor and the attractee.

2.9 American English <as of yet>

Although meaning can be affected by phonically based analogical attraction, as the examples in 2.6-2.8 show, the results of such analogical pressure need not always make sense. Rather, it can effect changes in the form alone even if aspects of the meaning are altered in unusual ways. A case in point is the expression *as of yet*, which seems to be an innovative crossing, a contamination that is, between two phrases, *as yet* and *as of now*, that were already present in the language. The emergence of *as of yet* means that either *as yet* has taken on *of* due to influence from *as of now*, or else *as of* has taken on *yet* as a possible complement due to influence from *as yet*. In either case, there is a shared phonic link through the word *as*, but there is as well a semantic link in that both are time expressions. Nonetheless, along with the analogical assimilation that leads one of these expressions in the direction of the other with regard to form, there is either a complication of or a shift in the semantics of the relevant pieces.

In particular, in the phrase *as of X*, the complement *X* generally has a definite and fixed time reference of some sort (e.g. *as of December*, *as of now*, *as of 3:33PM*, etc.); however, in the innovative *as of yet*, the complement has a very different kind of time reference, certainly not anything that could be characterized as definite in any sense, and thus a complication. Alternatively, one could say that the meaning of *yet* has shifted to accommodate its use in a new expression or that the requirements of *as of* have changed so as to allow a referentially vague term like *yet* as a complement. Either way there is a change beyond the new form, and the analogy leading to *as of yet*, with its phonic basis, is in large part responsible.

2.10 Latin <queō, nequeō>

The Latin verbs *queō* "I am able" and *nequeō* "I am not able" may well show morphological developments that under one account of their etymology would be a case of phonically based analogy. The standard etymology[10] treats *nequeō* as the older form, deriving from *neque* "and not" plus *eō* "go", originally in an impersonal passive formation *nequitur* "it does not go", with *queō* then a back formation created by slicing off the clear negative morpheme *ne-*. This suggestion fits the facts formally and may well be right, but it is not necessarily the most satisfying possibility on the semantic side. As an alternative, one might look to a different root as underlying these verbs, in particular

[10] See Ernout & Meillet (1939: s.v.).

Proto-Indo-European *k^{w}ey- "make, do" (as seen in Greek *poiéō*), so that the sense "be unable" would stem from "not to be done" (that is, "not doable"). While admittedly speculative (as is often the case with root etymologizing), in that case, these verbs would show 'assimilation' in their inflection to the form of *eō* "I go", in the following ways: from a preform 1PL.PRES *k^{w}ey-o-mos, one would expect either Latin *queumus** (with the phonetic development of *-eyo- seen in *ey-ont- "going" => *eunt-*) or *quēmus** (with the analogical development seen in forms like *monēmus* "we warn"). Instead what occurs for these verbs is *(ne)quīmus*, with the same root form as *īmus* "we go", from *ey-mos. Similarly, the infinitive is *(ne)quīre*, just like *īre* "to go", even though the expected outcome would be something like *(ne)quēre**. This cannot be proven conclusively, and it may well be that Ernout and Meillet are right in linking these verbs etymologically with *eō* from the start, but if the semantic connection is considered suspect, so that an alternative etymology is sought, then the later issue of how *(ne)queō* came to be linked with *eō* would have to be based not on their semantics but on the fact that they rhyme. That is, what would link the verbs, in this interpretation, and make the analogical influence possible, therefore, would be a phonic connection.

2.11 English <as far as…>

An example involving phonic analogy that affects syntax can be seen in the changes discussed by Rickford et al. 1995 with regard to the English construction beginning with *as far as* and signalling a focalized element. In particular, they note that a clear old construction in Modern English is that illustrated in (1), in which following the focalized element preceded by *as far as*, there is a verbal coda, usually *be concerned* though others such as *go* can also be found:

(1) a. As far as John is concerned, forget about him!
b. As far as John goes, forget about him!

In addition to this construction, there is another one, which Rickford et al. quite appropriately take to be innovative, in which *as far as* occurs but the verbal coda is lacking, as in (2):

(2) As far as John, forget about him!

Their main concern is the spread of the innovative construction in the past 200 years and especially in the later half of the 20th century, but they discuss various possible explanations for the appearance of the innovative pattern in the first place. One that they consider to be possible, but which in my view they pass over a bit too hastily (p. 115), is that given by Faris 1962 concerning possible involvement of another focalizing construction with *as for*, as in (3), which has no verbal coda:

(3) a. As for John, forget about him!
b. *As for John is concerned, forget about him!

The absence of the verbal coda in the *as for* construction would provide a model for its analogical absence in the *as far as* construction. But what is the basis for a connection between the two constructions? They are functionally linked, of course, in that both mark focused elements, but alongside this functional connection, there is another that cannot be ruled out, namely what Faris may have been hinting at when he referred to the influence

of 'the closely resembling *as for'* (p. 238): a phonic link. That is, one could claim that *as for* provided a suitable model for *as far as* based on the shared phonic form between the two of *as* and *f-V-r*. In this way, the innovative verbless construction would be a contamination or crossing (as seen above), an analogical creation with a phonic basis.

2.12 American English <being that>

My final example also is a case of syntax being affected by a phonically based analogy, and is quite parallel to the *as far as* example in 2.11. In this instance, the two older constructions that play a role, in my account, are the subordinate clauses (underlined) exemplified in (4):

(4) a. Seeing that John is here, we can start.
b. It being the case that John is here, we can start.

and the innovative construction is that illustrated in (5):

(5) Being that John is here, we can start.

All of these represent ways of stating the circumstances under which the action of the main clause occurs, (4a) with a gerund (or participle) that ostensibly is linked to the main clause subject and (4b) with an absolute construction containing an expletive *it* serving as subject to *being*. In the case of (5), there is as well a 'dangling' participle, in that *being* is not linked to any main clause argument, but also the syntactic anomaly of the suppression of the expletive subject of *being*, even though English in general is not a *pro*-Drop language.[11] How did the innovative construction in (5) arise? It is my contention that it is the result of a crossing of the two older constructions seen in (4), where the connection between the two is on the one hand functionally based in that both indicate attendant circumstances, but further, that it is aided by the phonic link between the two as well, essentially the rhyming of *seeing* with *being*, and thus due to the same sort of pressures that gave rise to the innovative *as far as* construction, and indeed the other innovative forms throughout section 2.

3 Conclusion

There is more that can be said about these examples and their collective effect. For instance, in some cases, the analogy results in a new form that is far from regular or simplified, far from 'optimal', as with the *being that* construction in 2.12, with its odd suppression of a pronoun that runs counter to otherwise quite general English subject requirements, or *as of yet* with its odd semantics or selectional anomaly. The suggestion that these anomalies emerge by analogical pressures means that analogy cannot be taken as an optimizing or regularizing force per se, except perhaps when applied to individual cases; that is, rather than leading to system-wide regularization and simplification ("system optimization" in the sense of Kiparsky 2000), analogy can introduce

[11] Admittedly, most treatments of *pro*-Drop refer to the suppression of subject pronouns in finite clauses (as in Modern Greek *tréxo* "I am-running" (literally, "am-running"). However, English has the free suppression of subject pronouns only in imperatives, and gerund/participial forms normally lack a subject only under circumstances of control from a main clause nominal (as in (4a)). Thus the absence of *it* here is innovative from a syntactic point of view.

complication into the system – the regularization would seem to be just on a very localized basis (in the sense of Joseph & Janda 1988), in that, as here, there is an 'inner logic', as it were, to the creation of *as of yet* because of the presence of *as yet* and *as of now*, at least in terms of its surface form; so also with *being that*.[12]

Finally, it must be emphasized that even though phonic effects, based on these examples, seem to be capable of playing an important role in establishing analogical links among forms, it is not the case that phonic effects hold sway every time one of these forms is uttered. Rather, as with any change, once a new form takes hold, the path by which it arrived at that particular shape is largely irrelevant. For instance, even though the currently widespread American English pronunciation of *often* with medial *–t-* has its origins in a spelling-based pronunciation, it is not the case that every time it is uttered now, speakers have the spelled form in mind inducing them into pronouncing the *–t-*; rather for most such speakers, *often* is simply learned with a *t* and thus always pronounced that way. So too in the examples discussed here: it is not the case that every innovative utterance of *academia* has a *macadamia* lurking behind it, so to speak. However, as the need to separate the impetus of an innovation from its spread is necessary in most accounts of language change in general, this aspect of the discussion merely places these examples in conformity with what is known about language change more generally.

References

Anderson, John; Hope C. Dawson; and Brian D. Joseph. 2010. Historical linguistics. *The Routledge linguistics encyclopedia* (3rd edition), ed. by Kirsten Malmkjaer, 225-251. London: Routledge Publishers.

Anttila, Raimo. 1972/1989. *An introduction to historical and comparative linguistics*. New York: Macmillan (2nd edn., *Historical and comparative linguistics*, Amsterdam: John Benjamins,1989).

Anttila, Raimo. 1977. *Analogy* (Trends in linguistics. State-of-the-art reports). The Hague: Mouton.

Anttila, Raimo. 2003. Analogy: The warp and woof of cognition. In Joseph and Janda, 425-440.

Anttila, Raimo and Warren Brewer. 1977. *Analogy: A basic bibliography***.** Amsterdam: John Benjamins Publishing Co.

Dawson, Hope C. 2005. *Morphological variation and change in the Rigveda: The case of* –au *vs.* –ā. Columbus: The Ohio State University Ph.D. Dissertation.

Ernout, Alfred and Antoine Meillet. 1939**.** *Dictionnaire étymologique de la langue latine. Histoire des mots*. Paris: Klincksieck**.**

Faris, Paul. 1962. 'As far as halfbacks, we're all right'. *American Speech* 37.236-38.

Haugen, Einar. 1969. *The Norwegian language in America. A study in bilingual behavior*. Bloomington: Indiana University Press.

Hock, Hans Henrich. 2003. Analogical change. In Joseph and Janda, 441-460.

Hock, Hans Henrich, and Brian D. Joseph. 1996. *Language history, language change, and language relationship. An introduction to historical and comparative linguistics*. Berlin: Mouton de Gruyter.

[12] This line of reasoning is pursued further in Joseph 2012.

Joseph, Brian D. 2008. Historical Linguistics in 2008: The state of the art. *Unity and Diversity of Languages*, ed. by Piet van Sterkenberg, 175-187. Amsterdam: John Benjamins Publishers.

Joseph, Brian D. 2012. Optimality, optimization, and analogy. Columbus: Ohio State University, ms.

Joseph, Brian D., and Richard D. Janda. 1988. The how and why of diachronic morphologization and demorphologization. *Theoretical morphology: Approaches in modern linguistics*, ed. by Michael Hammond and Michael Noonan, 193-210. San Diego: Academic Press.

Joseph, Brian D., and Richard D. Janda, eds. 2003. *The handbook of historical linguistics*. Oxford: Blackwell Publishers.

Kiparsky, Paul. 2000. Analogy as optimization: 'Exceptions' to Sievers' Law in Gothic. *Analogy, levelling, markedness. Principles of change in phonology and morphology*, ed. by Aditi Lahiri, 15-46. Berlin: Mouton de Gruyter.

Miller, D. Gary. 1982. *Homer and the Ionian epic tradition: Some phonic and phonological evidence against an Aeolic phase*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck.

Rickford, John R.; Thomas A. Wasow; Norma Mendoza-Denton; and Juli Espinoza. 1995. Syntactic variation and change in progress: Loss of the verbal coda in topic-restricting *as far as* constructions. *Language* 71.1.102-31.

Samuels, M. L. 1972. *Linguistic evolution; With special reference to English*. Cambridge: Cambridge University Press.

Vennemann, Theo. 1972. Phonetic analogy and conceptual analogy. *Schuchardt, the Neogrammarians, and the transformational theory of phonological change: Four essays by Hugo Schuchardt, Theo Vennemann, Terence H. Wilbur* (Linguistische Forschungen, 26), ed. by Theo Vennemann and Terence H. Wilbur, 181-204. Frankfurt am Main: Athenäum.

Watkins, C. 1995. *How to kill a dragon. Aspects of Indo-European poetics*. New York: Oxford University Press.

[joseph.1@osu.edu]

COORDINATION IN HYBRID TYPE-LOGICAL CATEGORIAL GRAMMAR

Yusuke Kubota and Robert Levine
University of Tokyo and Ohio State University

Abstract

We formulate explicit analyses of certain non-standard coordination examples discussed in Levine (2011) in a variant of categorial grammar called *Hybrid Type-Logical Categorial Grammar* (Kubota 2010; Kubota & Levine 2012; Kubota to appear). These examples are of theoretical importance since they pose significant challenges to the currently most explicit and most comprehensive analysis of coordination, formulated in a variant of HPSG called Linearization-based HPSG (Reape 1996; Kathol 1995) and advocated by various authors in the recent literature (Yatabe 2001; Crysmann 2003; Beavers & Sag 2004; Chaves 2007; Sag & Chaves 2008). This approach, which we call the Linearization-Based Ellipsis (LBE) approach to coordination, builds on the key idea that apparent non-standard coordination all reduce to constituent coordination under surface ellipsis. The seemingly heterogeneous set of data catalogued in Levine (2011), involving different types of non-standard coordination, uniformly point to an analysis in which the apparently incomplete constituents that are coordinated in the overt string are in fact complete (i.e. non-elliptical) constituents with full-fledged semantic interpretation, thus directly counterexemplifying the predictions of ellipsis-based approaches including the LBE variant. The sophisticated syntax-semantics interface of the framework

Yusuke Kubota and Robert Levine. Coordination in Hybrid TLCG. *OSUWPL* Volume 60, Spring 2013, pp. 21-50.

we propose in this paper straightforwardly captures the interactions between such non-standard coordination and various scopal expressions, demonstrating the real empirical payoff of the direct coordination analysis of non-standard coordination (of the kind widely adopted in categorial grammar) that has not been fully recognized in the previous literature.

1 Introduction

Levine (2011) provides a thorough critique of an approach to coordination that has become prominent in the recent literature of HPSG. The key idea of this approach, which we call the Linearization-Based Ellipsis (LBE) approach to coordination (Yatabe 2001; Crysmann 2003; Beavers & Sag 2004; Chaves 2007; Sag & Chaves 2008), is to analyze a wide range of coordination examples that apparently pose problems for phrase structure-based theories of syntax such as HPSG via a single mechanism of surface ellipsis. The analysis is technically implemented in a variant of HPSG that relaxes the mapping between the combinatoric structure and surface string known as Linearization-based HPSG (Reape 1996; Kathol 1995). In this paper, we take up some of the key examples from Levine (2011), which pose serious problems for the LBE approach and provide explicit analyses of them within a variant of categorial grammar called *Hybrid Type-Logical Categorial Grammar*. (Kubota 2010; Kubota & Levine 2012; Kubota to appear). While building heavily on ideas from previous literature of categorial grammar (CG), the proposed framework is novel in that it recognizes *both* the directionality-sensitive mode of implication (i.e. forward and backward slashes) familiar from the tradition of Lambek (1958) and the directionality-insensitive mode of implication employed in some of the more recent variants of CG pioneered by Oehrle (1994). This hybrid implication architecture enables a flexible and sophisticated syntax-semantics interface that is not available in previous variants of CG, and we show below that the analyses of the empirical phenomena taken up in this paper crucially exploits this flexibility and systematicity of the proposed framework. As will become clear below, due to the generality of the underlying logic, the present framework is not amenable to the kinds of criticisms that have been occasionally raised against previous variants of CG (especially CCG), which, for various reasons, do not entertain the flexibility of the logic-based syntax-semantics interface characteristic to CG in a fully general way.

The key hypothesis of the LBE approach to coordination, in its strongest version, is that a wide range of non-standard coordination such as the following (dependent cluster coordination (DCC) in (1a), right-node raising (RNR) in (1b), agreement anomaly in nominal head coordination in (1c) and unlike category coordination (UCC) in (1d)) can all be reduced to ordinary constituent coordination via surface ellipsis along the lines of (2).

(1) a. I gave Robin a book and Terry a pair of pliars.
b. I gave Robin, and Leslie offered Terry, a pair of pliars.
c. That man and woman are arguing again.
d. Robin is a Republican and proud of it.

(2) a. [$_S$ [$_S$ I gave Robin a book on Thursday] and [$_S$ ~~I gave~~ Leslie a book on Friday.]]
b. [$_S$ [$_S$ I gave Robin ~~a pair of pliars~~] and [$_S$ Leslie offered Terry, a pair of pliars]].
c. [$_{NP}$ [$_{NP}$ That man] and [$_{NP}$ ~~that~~ woman]] are arguing again.
d. [$_S$ [$_S$ Robin is a Republican] and [$_S$ ~~Robin is~~ proud of it.]]

The strikeout in (2) is meant to represent purely phonological deletion which is licensed on the condition that the same string appears in the other conjunct. This deletion operator (at least on the null hypothesis) is supposed to affect only the pronunciation of the sentence and not its semantic interpretation.

It should be noted that not all advocates of the LBE approach endorse an elliptical analysis for all of these cases (see, for example, Yatabe (2012), who expresses the view that an ellipsis-based approach is not appropriate for the latter two cases). However, since the plausibility of the hypothesis in part depends on its generality and since ellipsis-based analyses along the lines of (2) have been suggested for all of these phenomena by at least some authors advocating the LBE approach, we include them all here for completeness.

Following Levine (2011), we point out below that *none* of these cases are amenable to ellipsis-based analyses once we extend the dataset to cases in which the semantics of the coordinated expressions have non-trivial consequences for the compositional semantics of the whole sentence. The relevant examples involve the interactions between coordination and scopal expressions that appear outside the coordinate structure. We show below that the predictions of the LBE approach is systematically falsified in each such case. We then present explicit compositional analyses of these phenomena in Hybrid TLCG. We show that independently motivated analyses of each of these constructions interact properly with analyses of scopal elements in the proposed framework to yield the correct predictions in the relevant examples straightforwardly. Besides providing a general argument for a CG-based analysis of coordination, the data discussed in this paper thus provides strong empirical evidence for the proposed variant of categorial grammar (among other variants) in that the interactions between coordination and scopal expressions that they manifest call for exactly the sort of hybrid architecture of the syntax-semantics interface that is unique to it.

2 Contraindications for Linearization-Based Ellipsis (LBE)

2.1 Symmetrical, respective and summative predicates and nonconstituent coordination

Perhaps the strongest piece of evidence against ellipsis-based approaches to coordination comes from data such as (3a) and (4a), long known in the literature as a paradigm case which demonstrates the inadequacies of ellipsis-based analyses of nonconstituent coordination such as DCC and RNR (Abbott 1976; Jackendoff 1977; Gazdar 1981).

(3) a. I said $\left\{\begin{array}{l}\text{the same thing}\\ \text{different things}\end{array}\right\}$ to Robin on Thursday and (to) Leslie on Friday.

b. I said $\left\{\begin{array}{l}\text{the same thing}\\ \text{different things}\end{array}\right\}$ to Robin on Thursday and <u>(I) said $\left\{\begin{array}{l}\text{the same thing}\\ \text{different things}\end{array}\right\}$ to Leslie on Friday.</u>

(4) a. Robin reviewed, and Leslie read, $\left\{\begin{array}{l}\text{the same book}\\ \text{different books}\end{array}\right\}$.

b. Robin reviewed <u>$\left\{\begin{array}{l}\text{the same book}\\ \text{different books}\end{array}\right\}$</u>, and Leslie read $\left\{\begin{array}{l}\text{the same book}\\ \text{different books}\end{array}\right\}$.

Ellipsis-based analyses predict that the NCC examples in (3a) and (4b) are synonymous to the their constituent coordination counterparts in (3b) and (4b) (from which the NCC examples are derived by deleting the underlined parts—underlines in examples mean the same thing in all of the examples below), but this prediction is not borne out. (3a) and (4b) exhibit the so-called *internal reading* of *same* and *different* (Carlson 1987), which simply asserts the (non-)identity of the thing(s) in question. The constituent coordination counterparts in (3b) and (4b) have only the anaphoric, *external reading*, which presupposes the existence of some entity already salient in the discourse and asserts the (non-)identity between the things in question and that discourse-salient entity.

This type of non-parallel between NCC and their alleged clausal counterparts is not limited to symmetrical predicates, but is actually much more widespread. As noted, for example, by Abbott (1976) and Chaves (2012), essentially the same pattern is observed with the so-called 'respect readings' of sentences involving the adverb *respectively*, and the summative interpretations of numerical expressions such as *a total of $1000*. Examples of DCC are given in (5) and (6). Similar examples of RNR can be constructed easily.

(5) a. I lent *Barriers* and *Syntactic Structures* to Robin on Thursday and (to) Leslie on Friday, respectively.

b. I lent *Barriers* and *Syntactic Structures* to Robin on Thursday and <u>I lent *Barriers* and *Syntactic Structures* (to) Leslie on Friday, respectively.</u>

(6) a. I lent *$1000 in total* to Robin on Thursday and (to) Leslie on Friday.

b. I lent *$1000 in total* to Robin on Thursday and <u>I lent *$1000 in total*</u> (to) Leslie on Friday.

Here, too, the NCC examples have meanings that are not available in their alleged clausal counterparts, and this fact remains a mystery for ellipsis-based approaches to coordination.

It turns out that working out an explicit compositional semantics for this type of examples poses a significant challenge for any type of approach to NCC, be it ellipsis-based or not. In fact, as far as we are aware, except for Kubota (2010), there is no explicit proposal in the literature that provides a completely satisfactory solution for this problem. We

reproduce in section 3 Kubota's (2010) analysis of examples like (3a) and (4a) (involving symmetrical predicates), which exploits the hybrid implication architecture of the present framework.

2.2 UCC and extraction

The use of ellipsis to derive examples like (7a) from underlying 'source' structures like (7b) and thereby eliminate the 'unlikeness' of UCC was suggested by Beavers & Sag (2004) and then advocated more extensively by Chaves (2006).

(7) a. Robin is a Republican and proud of it.
b. Robin is a Republican and (Robin) is proud of it.

However, such an analysis leads to severe mispredictions once one considers more complex examples. Note first that strings like *rich and a Republican*, which exemplify the unlike category coordination, can be topicalized as in (8a):

(8) a. Rich and a Republican, Robin definitely is t.
b. Rich Robin definitely is t and a Republican Robin definitely is t.

The only way to derive (8a) from ellipsis is to assume an underlying source of the form in (8b) involving conjunction of full-fledged clauses. On this type of analysis, the following examples turn out to be crucially problematic:

(9) a. (Both) poor and a Republican, you can't possibly be t.
b. (Both) poor you can't possibly be t and a Republican you can't possibly be t.
(10) a. Dead drunk but in complete control of the situation, no one can be t.
b. Dead drunk no one can be t but in total control of the situation, no one can be t.

The ellipsis-based approach demands (9a) and (10a) to be derived from (9b) and (10b), but there is a mismatch in semantic interpretation that is essentially parallel to the symmetrical predicate data from the previous section. In (9a) and (10a), the modal scopes over the whole coordinated string *rich and a Republican*; in other words, in (9a), what is negated is the property of simultaneously being poor and a Republican. The non-elliptical sources in (9b) and (10a) lack that interpretation totally; they can only be interpreted as a conjunction of negation, which has a stricter truth conditions than their alleged elided counterparts.

The same observation can be replicated in yet another displacement construction, namely, pseudocleft. In (11a), a conjoined unlike category occupies the focus position of a

pseudocleft sentence. On the ellipsis-based analysis, this example has to be derived from the underlying source in (11b), but again, there is a semantic mismatch between the UCC example and its alleged underlying source.

(11) a. What you cannot become (simultaneously) is highly intelligent and (yet) a raving fundamentalist.
b. What you cannot (simultaneously) become is highly intelligent and (yet) what you cannot (simultaneously) become is a raving fundamentalist.

2.3 Nominal head coordination under a singular determiner

Finally, we consider the case of nominal head coordination under a singular determiner exemplified by data such as (12):

(12) That man and woman are arguing again.

Chaves (2007) and Sag & Chaves (2008) suggest the possibility of deriving (12) from an underlying source of the form of (13). An apparent advantage of such an analysis is that it provides an immediate (and simple) solution for the seemingly anomalous agreement pattern in (12) (where a singular determiner is used for an NP which clearly refers to multiple individuals).

(13) That man and that woman are arguing again.

Consideration of a wider range of data, however, once again reveals that such an ellipsis-based analysis is too simplistic. Examples like the following noted by Heycock & Zamparelli (2005) in which a symmetrical modifier appears in the the 'ellipsis' environment resist an analysis along the lines of (13).

(14) a. That ill-matched man and woman are fighting again.
b. *That ill matched man and that ill-matched woman are fighting again.

(15) a. That mutually hostile judge and defense attorney were constantly sniping at each other during the trial.
b. *That mutually hostile judge and that mutually hostile defense attorney were constantly sniping at each other during the trial.

The alleged underlying sources for (14a) and (15a), given in (14b) and (15b), are simply ungrammatical. Here again, the problem essentially stems from the fact that an ellipsis-based analysis gets the semantic scope wrong. For example, to derive the right meaning for

(14a), the symmetrical predicate *ill-matched* has to scope over the conjoined noun *man and woman* so as to establish the right relation between the man and the woman in question. Such an interpretation cannot be obtained from the 'source' structure (14b), where two distinct tokens of *ill-matched* modify the nouns *man* and *woman* separately within each conjunct.

2.4 Summary

The LBE approach to coordination at first sight appears to provide a simple and uniform solution for a wide range of apparently heterogeneous set of non-standard coordination phenomena. However, as we have seen above, the success is illusory. Once we look beyond the simplest cases exemplified by (1), the ellipsis-based approach faces several severe difficulties. Specifically, in all of the cases discussed in this section, whose key examples are repeated in (the a.-examples of) (16)–(18), we see a systematic interaction between the coordinated expression and scopal operators that appear *outside* the coordinate structure in the overt string.

(16) a. I said { the same thing / different things } to Robin on Thursday and (to) Leslie on Friday.

b. I said { the same thing / different things } to Robin on Thursday and ~~(I) said~~ { **~~the same thing~~** / **~~different things~~** } ~~to~~ Leslie on Friday.

(17) a. (Both) poor and a Republican, you can't possibly be t.

b. (Both) poor ~~you **can't possibly** be~~ t and a Republican you can't possibly be t.

(18) a. That ill-matched man and woman are fighting again.

b. That ill-matched man and ~~that **ill-matched**~~ woman are fighting again.

In all these cases, the observed empirical pattern is that the operator scopes over the whole coordinate structure in a way that mirrors the surface form of the sentence. The LBE approach systematically mispredicts in such cases, since, on the ellipsis-based analysis, the scopal operator is part of the elided string (as in b. above) and hence appears *inside* each conjunct, in effect reversing the scopal relation between the operator and the coordinate structure from what is actually observed.

3 Hybrid Type-Logical Categorial Grammar

The central characteristic of the variant of CG that we propose in this paper is that it recognizes two kinds of implication, namely, the order-sensitive forward and backward slashes familiar from the tradition of Type-Logical Categorial Grammar originating from Lambek's (1958) work, and the order-insensitive mode of implication tied to phonological λ-binding in more recent variants of CG (stemming from Oehrle's (1994) work) that

relegate word order-related information from the combinatoric component of syntax to a separate morpho-phonological component. As will become clear below, the hybrid architecture of the present framework is exploited crucially in capturing the interactions between coordination and various scopal expressions. Directional variants of categorial grammar provide an elegant analysis of non-standard coordination (especially nonconstituent coordination including both DCC and RNR), while they are subotimal for scopal phenomena due to the inherently directional nature of the underlying calculus. By contrast, variants of CG that relegate word order entirely to a separate prosodic component enables a straightforward treatment of scopal phenomena, but they have the drawback that the elegant analysis of (non-standard) coordination in directional variants of CG is lost, due to the fact that syntactic categories of linguistic expressions do not carry order-related information (an aspect of directional variants of CG that is crucially exploited in the analysis of coordination). Hybrid TLCG entertains the advantages of both directional and non-directional variants of CG, by recognizing both kinds of implication within a single calculus. The complex interactions between coordination and scopal expressions exhibited by the data observed in section 2 requires exactly this kind of architecture, providing empirical evidence for the novel architecture of CG embodied in the proposed framework.

3.1 Hybrid Implication System as an Underlying Logic

Following Oehrle (1994), we write linguistic expressions as tuples of phonological form, semantic interpretation and syntactic category (written in that order). Our system recognizes both directional modes of implication ($/$ and $\backslash$) and a non-directional mode of implication that we call the *vertical slash* ($|$, for which we write the argument to its right, just as with $/$). The full set of inference rules posited in the calculus, consisting of the Introduction and Elimination rules for the three kinds of slashes, are given in (19).

(19)

Connective	Introduction	Elimination
$/$	$\dfrac{\begin{array}{c}\vdots\ [\varphi; x; A]^n\ \vdots \\ \vdots\quad\ \vdots\quad\ \vdots\end{array}}{\dfrac{b \circ \varphi; \boldsymbol{f}; B}{b; \lambda x.\boldsymbol{f}; B/A}\,/\mathrm{I}^n}$	$\dfrac{a; \boldsymbol{f}; A/B \quad b; \boldsymbol{g}; B}{a \circ b; \boldsymbol{f}(\boldsymbol{g}); A}\,/\mathrm{E}$
$\backslash$	$\dfrac{\begin{array}{c}\vdots\ [\varphi; x; A]^n\ \vdots \\ \vdots\quad\ \vdots\quad\ \vdots\end{array}}{\dfrac{\varphi \circ b; \boldsymbol{f}; B}{b; \lambda x.\boldsymbol{f}; A\backslash B}\,\backslash\mathrm{I}^n}$	$\dfrac{b; \boldsymbol{g}; B \quad a; \boldsymbol{f}; B\backslash A}{b \circ a; \boldsymbol{f}(\boldsymbol{g}); A}\,\backslash\mathrm{E}$
$\lvert$	$\dfrac{\begin{array}{c}\vdots\ [\varphi; x; A]^n\ \vdots \\ \vdots\quad\ \vdots\quad\ \vdots\end{array}}{\dfrac{b; \boldsymbol{f}; B}{\lambda\varphi.b; \lambda x.\boldsymbol{f}; B\lvert A}\,\lvert\mathrm{I}^n}$	$\dfrac{a; \boldsymbol{f}; A\lvert B \quad b; \boldsymbol{g}; B}{a(b); \boldsymbol{f}(\boldsymbol{g}); A}\,\lvert\mathrm{E}$

The key difference between the directional slashes ($/$ and $\backslash$) and the non-directional slash ($|$) is that while the Introduction and Elimination rules for the former refer to the phonological forms of the input and output strings (so that, for example, the applicability of the /I rule is conditioned on the presence of the phonology of the hypothesis φ on the right periphery of the phonology of the input $b \circ \varphi$),[1] the rules for the latter are not constrained that way. For reasoning involving $|$, the phonological terms themselves fully specify the ways in which the output phonology is constructed from the input phonologies. Specifically, for $|$, the phonological operations associated with the Introduction and Elimination rules mirror exactly the semantic operations for these rules: function application and λ-abstraction, respectively. We assume that the binary connective $\circ$ in the phonological term calculus represents the string concatenation operation and that $\circ$ is associative in both directions. For notational convenience, we implicitly assume the associativity axiom $(\varphi_1 \circ \varphi_2) \circ \varphi_3 \equiv \varphi_1 \circ (\varphi_2 \circ \varphi_3)$ and leave out all the brackets indicating the internal constituency of complex phonological terms.[2] The phonological term calculus is a lambda calculus, and we also take the equivalence between β-reduced and unreduced terms to be an axiom.[3]

It should be clear from the way the above rules are formulated that the present system without the rules for $|$ is equivalent to the Lambek calculus (Lambek 1958), while the system with only the rules for $|$ is essentially equivalent to the term-labelled calculus of Oehrle (1994), Lambda Grammar (Muskens 2003), Abstract Categorial Grammar (de Groote 2001), and Linear Grammar (Pollard 2011), with some irrelevant implementational details aside.

3.2 Basic Analyses of Scope and Coordination

As was first demonstrated by Oehrle (1994), λ-binding in the phonological component enables an insightful analysis of quantifier scope, a problem whose general solution has turned out to pose significant theoretical challenges to directional variants of

[1] In this respect, the present calculus follows most closely Morrill & Solias (1993) and Morrill (1994); see Moortgat (1997) and Bernardi (2002) for an alternative formulation where sensitivity to directionality is mediated through a presumed correspondence between surface string and the form of structured antecedents in the sequent-style notation of natural deduction.

[2] For a more fine-grained control of surface morpho-phonological constituency, see Kubota & Pollard (2010) (and also Muskens (2007) for a related approach), which formalizes the notion of multi-modality from the earlier TLCG literature (Moortgat & Oehrle 1994; Morrill 1994) by modelling the mapping from syntax to phonology by means of an interpretation of (phonological) λ-terms into preorders.

[3] Note that the equivalence enforced by the associativity axiom is a property of the prosodic calculus directly reflecting the structures of the objects that the prosodic terms are supposed to model whereas equivalence under β-reduction is a formal property of the calculus itself. In this sense, the two equivalence relations we implicitly assume here are of somewhat different nature. In particular, the latter assumption underlies the fundamentally hybrid nature of the proposed system of tripartite inference as a whole in that β-reductions of prosodic terms that result from inferences involving the vertical slash sometimes play a crucial role for the applicability of subsequent inferences involving the directional slashes. A deductive system with such a radically hybrid property is unheard of and is surely unorthodox, and its formal underpinnings need to be investigated more closely, but we leave this task for future study.

CG.[4] As will become clear below, Oehrle's analysis illuminates the fundamental nature of scope-taking expressions by capturing the tight correlation between λ-binding in semantics and the corresponding binding of the 'gap' in the phonology, which is merely enforced by a bunch of coincidental stipulations in Montague's syncategorematic treatment of quantifying-in (which in turn is modelled in various ways in Lambek calculus-based variants of TLCG essentially carrying over the stipulative nature of Montague's solution as is). Unlike Montague's original treatment, in Oehrle's (1994) analysis the mismatch between linear position and scopal domain is a predicted *consequence* of hypothetical reasoning in the underlying logic.

The following analysis of the inverse-scope ($\forall > \exists$) reading of the sentence *Someone talked to everyone yesterday* illustrates how this works. We first hypothetically assume NPs in the surface positions in which the quantifiers appear. After the whole sentence is built up, withdrawing one of the hypotheses and binding the variables (in both phonology and semantics) via Vertical Slash Introduction produces a lambda-abstracted meaning/phonology pair that can be given to a quantifier as an argument. Combining the quantifier with such a lambda abstract has the effect that semantically the quantifier scopes over the whole expression but phonologically the string of the quantifier is inserted to the variable slot that is explicitly kept track of in the sentence's phonology via phonological lambda binding. The order in which the quantifiers are introduced in the derivation corresponds to their relative scope. Thus, in (20), since the universal quantifier in the object position is introduced in the derivation after the subject position quantifier, the inverse-scope interpretation is derived.

(20)

$$\dfrac{\lambda\sigma.\sigma(\mathsf{everyone});\ \forall_{\mathbf{person}};\ \mathrm{S}|(\mathrm{S}|\mathrm{NP}) \qquad \dfrac{\lambda\sigma.\sigma(\mathsf{someone});\ \exists_{\mathbf{person}};\ \mathrm{S}|(\mathrm{S}|\mathrm{NP}) \qquad \dfrac{\dfrac{\dfrac{\mathsf{talked}\circ\mathsf{to};\ \mathbf{talk\text{-}to};\ (\mathrm{NP}\backslash\mathrm{S})/\mathrm{NP} \qquad \left[\begin{smallmatrix}\varphi_1;\\ x;\mathrm{NP}\end{smallmatrix}\right]^1}{\mathsf{talked}\circ\mathsf{to}\circ\varphi_1;\ \mathbf{talk\text{-}to}(x);\mathrm{NP}\backslash\mathrm{S}}/\mathrm{E} \qquad \left[\begin{smallmatrix}\varphi_2;\\ y;\mathrm{NP}\end{smallmatrix}\right]^2}{\varphi_2\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1;\ \mathbf{talk\text{-}to}(x)(y);\mathrm{S}}\backslash\mathrm{E} \qquad \mathsf{yesterday};\ \mathbf{yest};\mathrm{S}\backslash\mathrm{S}}{\varphi_2\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1\circ\mathsf{yesterday};\ \mathbf{yest}(\mathbf{talk\text{-}to}(x)(y));\mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi_2.\varphi_2\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1\circ\mathsf{yesterday};\ \lambda y.\mathbf{yest}(\mathbf{talk\text{-}to}(x)(y));\mathrm{S}|\mathrm{NP}}|\mathrm{I}^2}{\mathsf{someone}\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1\circ\mathsf{yesterday};\ \exists_{\mathbf{person}}(\lambda y.\mathbf{yest}(\mathbf{talk\text{-}to}(x)(y)));\mathrm{S}}|\mathrm{E}}{\lambda\varphi_1.\mathsf{someone}\circ\mathsf{talked}\circ\mathsf{to}\circ\varphi_1\circ\mathsf{yesterday};\ \lambda x.\exists_{\mathbf{person}}(\lambda y.\mathbf{yest}(\mathbf{talk\text{-}to}(x)(y)));\mathrm{S}|\mathrm{NP}}|\mathrm{I}^1}{\mathsf{someone}\circ\mathsf{talked}\circ\mathsf{to}\circ\mathsf{everyone}\circ\mathsf{yesterday};\ \forall_{\mathbf{person}}(\lambda x.\exists_{\mathbf{person}}(\lambda y.\mathbf{yest}(\mathbf{talk\text{-}to}(x)(y))));\mathrm{S}}|\mathrm{E}$$

As we have seen above, the use of the non-directional mode of implication enables a perspicuous treatment of scope, which essentially involves hypothetical reasoning

[4]Independent motivation for this technique comes from an analysis of extraction (see section 4.2 below) and Gapping (Kubota & Levine 2012).

for which the information of word order is in principle irrelevant. However, in a system with the non-directional mode of implication alone (such as Oehrle's (1994) original term-labelled calculus as well as Lambda Grammar (Muskens 2003), Abstract Categorial Grammar (de Groote 2001) and Linear Grammar (Pollard 2011)), coordination turns out to pose a significant challenge. This point was already noted by Muskens (2001) (see also Kubota (2010:section 3.2.1) for a more extended discussion on this point), but no satisfactory solution has been proposed for this problem in the literature until the present. The problem essentially comes from the fact that such frameworks recognize linguistic expressions with functional phonologies, and semantically functional expressions such as verbs come with functional phonologies which specify the way in which the phonologies of the arguments are combined with the phonology provided by the verb. However, the standard analysis of coordination in CG rests on the assumption that the phonologies of all linguistic expressions are simply strings (and not functions over strings). Thus, even the simplest cases of coordination involving lexical verbs such as *John saw and kissed Mary* turns out to be extremely problematic for such approaches. To analyze the coordination of expressions with functional phonologies (with different arity), we need to posit a series of prosodically distinct entries of *and* (and *or*). Working this out technically is by no means a trivial issue, but the more serious problem is that non-directional variants of CG do not distinguish between left-looking from right-looking functors. Thus, with an entry for *and* which coordinates type $str \rightarrow str$ conjuncts (which is needed for the analysis of Right-Node Raising examples like *John saw, and Bill hit, Peter*), an S|NP expression *John saw* __ will be able to be coordinated with an expression __ *hit Peter* of the same type, resulting in quite serious overgeneration.

Since an analysis of (non-standard) coordination is central to our account of the interaction between coordination and scope, rather than attempting to solve this recalcitrant theoretical problem within a non-directional variant of CG, we propose to extend such a system by recognizing both the directional and non-directional modes of implication within a single calculus, incorporating the insights of the Lambek-inspired tradition of TLCG. Since our calculus recognizes the Introduction and Elimination rules for forward and backward slashes from standard TLCG, the analysis of NCC in it straightforwardly carries over to the present setup, which we illustrate below.

With the Introduction and Elimination rules for directional slashes, the analysis of nonconstituent coordination originally due to Dowty's (1988) and Steedman's (1985) CCG analyses and later incorporated in TLCG by Morrill (1994) carries over to the present setup straightforwardly. The idea behind this analysis is essentially that, in the setup of TLCG, hypothetical reasoning with forward and backward slashes enables us to reanalyze any substring of a sentence as a full-fledged 'constituent' (with an appropriate, higher-order semantic interpretation) that has the right combinatorial property such that it returns a sentence when it combines with the rest of the sentence. The derivation in (22) shows how the string *Bill the book* in (21) is reanalyzed as such a non-standard constituent.

(21) Mary gave [Bill the book] and [John the record].

(22)

$$\dfrac{\dfrac{\dfrac{[\varphi;\ f;\ \mathrm{VP/NP/NP}\ ]^1 \quad \mathsf{bill};\ \mathbf{b};\ \mathrm{NP}}{\varphi \circ \mathsf{bill};\ f(\mathbf{b});\ \mathrm{VP/NP}}\,/\mathrm{E} \quad \mathsf{the} \circ \mathsf{book};\ \textbf{the-book};\ \mathrm{NP}}{\varphi \circ \mathsf{bill} \circ \mathsf{the} \circ \mathsf{book};\ f(\mathbf{b})(\textbf{the-book});\ \mathrm{VP}}\,/\mathrm{E}}{\mathsf{bill} \circ \mathsf{the} \circ \mathsf{book};\ \lambda f.f(\mathbf{b})(\textbf{the-book});\ (\mathrm{VP/NP/NP})\backslash\mathrm{VP}}\,\backslash\mathrm{I}^1$$

The key step in the above derivation is the hypothetical assumption of a ditransitive verb. This hypothetical verb combines with the two object NPs *Bill* and *the book* just like ordinary ditransitive verbs and forms a VP. Then, the hypothesis is withdrawn to assign the category $(\mathrm{VP/NP/NP})\backslash\mathrm{VP}$ to the string *Bill the book*. Intuitively, this is saying that this string is something that becomes a VP if it finds a ditransitive verb to its left. Once this complex category is assigned to the string *Bill the book*, the rest just involves coordinating this non-standard constituent with another constituent with the same syntactic category via the standard generalized conjunction category for the coordinator *and* (where $\sqcap$ denotes generalized conjunction *a la* Partee & Rooth (1983)), and then putting the whole coordinated expression together with the verb and the subject NP as in (23).

(23)

$$\dfrac{\mathsf{mary};\ \mathbf{m};\ \mathrm{NP} \qquad \dfrac{\mathsf{gave};\ \textbf{give};\ \mathrm{VP/NP/NP} \qquad \dfrac{\mathsf{bill} \circ \mathsf{the} \circ \mathsf{book};\ \lambda f.f(\mathbf{b})(\textbf{the-book});\ (\mathrm{VP/NP/NP})\backslash\mathrm{VP} \qquad \dfrac{\mathsf{and};\ \lambda V\lambda W.W \sqcap V;\ (X\backslash X)/X \qquad \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \lambda f.f(\mathbf{j})(\textbf{the-record});\ (\mathrm{VP/NP/NP})\backslash\mathrm{VP}}{\mathsf{and} \circ \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \lambda W.W \sqcap \lambda f.f(\mathbf{j})(\textbf{the-record});\ ((\mathrm{VP/NP/NP})\backslash\mathrm{VP})\backslash((\mathrm{VP/NP/NP})\backslash\mathrm{VP})}\,/\mathrm{E}}{\mathsf{bill} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \lambda f.f(\mathbf{b})(\textbf{the-book}) \sqcap \lambda f.f(\mathbf{j})(\textbf{the-record});\ (\mathrm{VP/NP/NP})\backslash\mathrm{VP}}\,\backslash\mathrm{E}}{\mathsf{gave} \circ \mathsf{bill} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \textbf{give}(\mathbf{b})(\textbf{the-book}) \sqcap \textbf{give}(\mathbf{j})(\textbf{the-record});\ \mathrm{VP}}\,\backslash\mathrm{E}}{\mathsf{mary} \circ \mathsf{gave} \circ \mathsf{bill} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{john} \circ \mathsf{the} \circ \mathsf{record};\ \textbf{give}(\mathbf{b})(\textbf{the-book})(\mathbf{m}) \wedge \textbf{give}(\mathbf{j})(\textbf{the-record})(\mathbf{m});\ \mathrm{S}}\,\backslash\mathrm{E}$$

Thus, positing both directional and non-directional modes of implication within a single calculus enables straightforward analyses of two kinds of major empirical phenomena (i.e. coordination and scopal expressions) that pose problems for previous variants of CG, which recognizes only one of these two types of implication. But the real strength of the hybrid implication architecture of the present framework becomes fully apparent in the analyses of phenomena like those discussed in section 2 in which coordination interacts with scopal expressions. These phenomena call for a system in which the mechanisms dealing with word order-related inferences (for coordination) and those dealing with order-insensitive reasoning (for scope) interact with one another systematically. The present framework provides precisely such an architecture, and we will see in the next section that the proper analysis of these more complex cases in fact falls out straightforwardly from the hybrid architecture of the present framework.

4 Coordination in Hybrid Type-Logical Categorial Grammar

4.1 NCC and symmetrical predicates

For the analysis of symmetrical predicates, we adopt the proposal by Barker (2007) in terms of *parasitic scope*. Pollard (2009) (described in Pollard & Smith (to appear)) implements this analysis in a term-labelling system with phonological λ-abstraction like Oehrle's (1994) setup. We adopt this implementation in our account of the interaction between NCC and symmetrical predicates. Barker's analysis of symmetrical predicates involves the following three elements as the key components in the semantic analysis of symmetrical predicates such as *(the) same*:[5]

(i) a property provided by the 'head noun' modified by the symmetrical predicate

(ii) a sum-denoting expression

(iii) a relation provided by the rest of the sentence (i.e. a structure obtained by abstracting over the NP containing the symmetrical predicate and the sum-denoting expression from the whole sentence)

For an example like (24), (i)–(iii) above are instantiated by the noun *waiter*, the coordinated NP *John and Bill*, and the transitive verb *served*, respectively.

(24) The same waiter served John and Bill.

The semantic contribution of *(the) same* on the internal reading is to assert the existence of some unique waiter (i.e. an individual satisfying the property (i)) such that the x-served-y relation (i.e. the relation provided by (iii)) holds between that individual and each atomic subpart of the plurality of John and Bill (i.e. the sum denoted by (ii)). Technically, the denotation of the symmetrical predicate *the same* is formulated as in (25), as a relation between its three semantic arguments (i)–(iii):

[5] We adopt this three-place function analysis of *same* due to Barker (2007) here for expository purposes. Note, however, that this analysis does not easily generalize to cases involving multiple occurrences of *same* invoking the internal reading with respect to the same plural entity at the same time, as exemplified by the following example:

(i) John and Bill bought the same book at the same store (on the same day ...).

In Kubota & Levine (2013), we provide a more complete analysis of symmetrical predicates which can deal with iterated *same* examples like (i). This latter analysis is superior to the one we adopt here also in that it extends straightforwardly to related expressions such as 'respective' and summative predicates observed above in (5) and (6), and captures the complex (yet systematic) interactions between these three classes of phenomena in a uniform manner.

(25) $\lambda P \lambda Q \lambda X.\exists y[P(y) \wedge \forall x.x <_a X \rightarrow Q(y)(x)]$

Syntactically, the key steps of the derivation involves forming constituents that respectively provide the three semantic arguments to the symmetrical predicate and combining them in the right order. Specifically, the relation-denoting expression can be obtained by abstracting over the positions corresponding to the NP containing *same* and the sum-denoting expression via successive applications of the Vertical Slash introduction rule (cf. the steps down to the fifth line of the rightmost chunk in (26)). This creates a doubly abstracted proposition of syntactic category $(\mathrm{S}|\mathrm{NP})|\mathrm{NP}$ and this doubly abstracted proposition is then given as an argument to the expression *the same*, together with the other two arguments. Crucially, the phonology of the symmetrical predicate is specified as a higher-order functional phonological term which, if it takes the three arguments it semantically requires in the right order, produces the string of words corresponding to the surface form of the sentence.

(26)

$$\dfrac{\mathsf{served};\ \mathbf{serve};\ \mathrm{VP/NP} \qquad \left[\begin{array}{l}\varphi;\\ x;\mathrm{NP}\end{array}\right]^1}{\mathsf{served} \circ \varphi;\ \mathbf{serve}(x);\ \mathrm{VP}}\ /\mathrm{E}$$

$$\dfrac{\left[\begin{array}{l}\psi;\\ y;\mathrm{NP}\end{array}\right]^2 \qquad \mathsf{served} \circ \varphi;\ \mathbf{serve}(x);\ \mathrm{VP}}{\psi \circ \mathsf{served} \circ \varphi;\ \mathbf{serve}(x)(y);\ \mathrm{S}}\ \backslash\mathrm{E}$$

$$\dfrac{\psi \circ \mathsf{served} \circ \varphi;\ \mathbf{serve}(x)(y);\ \mathrm{S}}{\lambda\varphi.\psi \circ \mathsf{served} \circ \varphi;\ \lambda x.\mathbf{serve}(x)(y);\ \mathrm{S}|\mathrm{NP}}\ |\mathrm{I}^1$$

$$\dfrac{\lambda\varphi.\psi \circ \mathsf{served} \circ \varphi;\ \lambda x.\mathbf{serve}(x)(y);\ \mathrm{S}|\mathrm{NP}}{\lambda\psi\lambda\varphi.\psi \circ \mathsf{served} \circ \varphi;\ \lambda x.\mathbf{serve}(x)(y);\ \mathrm{S}|\mathrm{NP}|\mathrm{NP}}\ |\mathrm{I}^2$$

$$\dfrac{\lambda\psi\lambda\sigma.\sigma(\mathsf{the} \circ \mathsf{same} \circ \psi);\ \mathbf{same};\ ((\mathrm{S}|X)|((\mathrm{S}|X)|\mathrm{NP}))|\mathrm{N} \qquad \mathsf{waiter};\ \mathbf{waiter};\ \mathrm{N}}{\lambda\sigma.\sigma(\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter});\ \mathbf{same}(\mathbf{waiter});\ (\mathrm{S}|X)|((\mathrm{S}|X)|\mathrm{NP})}\ |\mathrm{E}$$

$$\dfrac{\lambda\sigma.\sigma(\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter});\ \mathbf{same}(\mathbf{waiter});\ (\mathrm{S}|X)|((\mathrm{S}|X)|\mathrm{NP}) \qquad \lambda\psi\lambda\varphi.\psi \circ \mathsf{served} \circ \varphi;\ \lambda x.\mathbf{serve}(x)(y);\ \mathrm{S}|\mathrm{NP}|\mathrm{NP}}{\lambda\varphi.\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter} \circ \mathsf{served} \circ \varphi;\ \mathbf{same}(\mathbf{waiter})(\lambda y \lambda x.\mathbf{serve}(x)(y));\ \mathrm{S}|\mathrm{NP}}\ |\mathrm{E}$$

$$\dfrac{\begin{array}{c}\vdots \quad \vdots\\ \mathsf{john} \circ \mathsf{and} \circ \mathsf{bill};\ \mathbf{j} \oplus \mathbf{b};\ \mathrm{NP}\end{array} \qquad \lambda\varphi.\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter} \circ \mathsf{served} \circ \varphi;\ \mathbf{same}(\mathbf{waiter})(\lambda y \lambda x.\mathbf{serve}(x)(y));\ \mathrm{S}|\mathrm{NP}}{\mathsf{the} \circ \mathsf{same} \circ \mathsf{waiter} \circ \mathsf{served} \circ \mathsf{john} \circ \mathsf{and} \circ \mathsf{bill};\ \mathbf{same}(\mathbf{waiter})(\lambda y \lambda x.\mathbf{serve}(x)(y))(\mathbf{j} \oplus \mathbf{b});\ \mathrm{S}}\ |\mathrm{E}$$

(In the above derivation, the variables abstracted over happen to correspond to subject and object arguments of the transitive verb, and thus these two abstraction steps might appear to be (semantically) superfluous. However, as will become clear in more complicated derivations below, this is not generally the case, and we need the two abstraction steps for the vertical slash in order to provide a general treatment of symmetrical predicates.) By unpacking the final translation for the sentence derived in (26), we get the following truth conditions:

(27) $\mathbf{same}(\mathbf{waiter})(\lambda y \lambda x.\mathbf{serve}(x)(y))(\mathbf{j} \oplus \mathbf{b}) = \exists y[\mathbf{waiter}(y) \wedge \forall x.x <_a \mathbf{j} \oplus \mathbf{b} \rightarrow \mathbf{serve}(x)(y)]$

This says that there is a waiter such that for both of John and Bill, it is the case that that waiter served him. In other words, there is a unique waiter who served both John and Bill. This correctly captures the meaning of the sentence on the relevant reading.

This analysis of symmetrical predicates by Barker implemented in a system with phonological lambda abstraction interacts straightforwardly with the direct licensing analysis of NCC from directional CG to assign the right interpretations for sentences like (28).

(28) Terry said the same thing to Robin on Thursday and to Leslie on Friday.

The crucial assumption that enables a straightforward extension of the analysis of symmetrical predicates for the simpler case involving coordination of simple NPs (denoting sums of type e objects) above to more complex cases like (28) is that in the lexical entry for *the same* in (27), the type of the sum-denoting expression (and, correspondingly, of the relation that takes subparts of that sum as one of its arguments) is polymorphic. Specifically, in the case of (28), the sum involved is a sum of higher-order semantic objects of type $e \to (e \to e \to e \to t) \to e \to t$. Other than this slight complication in the semantic type, the function of the symmetrical predicate *the same* is the same as in the previous case: it asserts that an identical relation holds between each subpart of this sum and some unique entity satisfying the descriptive content provided by the nominal head that the symmetrical predicate combines with. Thus, the analysis is essentially parallel to the simpler case involving coordination of ordinary NPs in (24). The derivation for (28) is given in (29).

(29)

$$\dfrac{\dfrac{\dfrac{\begin{matrix}\text{terry};\\ \mathbf{t}; \mathrm{NP}\end{matrix} \quad \dfrac{\begin{matrix}\text{said};\\ \mathbf{say}; \mathrm{VP/PP/NP}\end{matrix} \quad \dfrac{[\varphi_1; x; \mathrm{NP}]^1 \quad [\varphi_2; f; \mathrm{NP\backslash(VP/PP/NP)\backslash VP}]^2}{\varphi_1 \circ \varphi_2; f(x); \mathrm{(VP/PP/NP)\backslash VP}}\backslash\mathrm{E}}{\text{said} \circ \varphi_1 \circ \varphi_2; f(x)(\mathbf{say}); \mathrm{VP}}\backslash\mathrm{E}}{\text{terry} \circ \text{said} \circ \varphi_1 \circ \varphi_2; f(x)(\mathbf{say})(\mathbf{t}); \mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi_2.\text{terry} \circ \text{said} \circ \varphi_1 \circ \varphi_2; \lambda x.f(x)(\mathbf{say})(\mathbf{t}); \mathrm{S|(NP\backslash(VP/PP/NP)\backslash VP)}}|\mathrm{I}^1}{\lambda\varphi_2\lambda\varphi_1.\text{terry} \circ \text{said} \circ \varphi_1 \circ \varphi_2; \lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t}); \mathrm{S|(NP\backslash(VP/PP/NP)\backslash VP)|NP}}|\mathrm{I}^2$$

$$\dfrac{\begin{matrix}\vdots \quad \vdots\\ \text{to} \circ \text{robin} \circ \text{on} \circ \text{thursday};\\ \lambda x\lambda P.\mathbf{onTh}(P(x)(\mathbf{r}));\\ \mathrm{NP\backslash(VP/PP/NP)\backslash VP}\end{matrix} \quad \dfrac{\begin{matrix}\text{and};\\ \lambda X\lambda Y.X \oplus Y;\\ (X\backslash X)/X\end{matrix} \quad \begin{matrix}\vdots \quad \vdots\\ \text{to} \circ \text{leslie} \circ \text{on} \circ \text{friday};\\ \lambda x\lambda P.\mathbf{onFr}(P(x)(\mathbf{l}));\\ \mathrm{NP\backslash(VP/PP/NP)\backslash VP}\end{matrix}}{\begin{matrix}\text{and} \circ \text{to} \circ \text{robin} \circ \text{on} \circ \text{thursday};\\ \lambda X.X \oplus [\lambda x\lambda P.\mathbf{onFr}(P(x)(\mathbf{l}))];\\ \mathrm{(NP\backslash(VP/PP/NP)\backslash VP)\backslash(NP\backslash(VP/PP/NP)\backslash VP)}\end{matrix}}/\mathrm{E}}{\begin{matrix}\text{to} \circ \text{robin} \circ \text{on} \circ \text{thursday} \circ \text{and} \circ \text{to} \circ \text{leslie} \circ \text{on} \circ \text{friday};\\ \lambda x\lambda P\mathbf{onFr}(P(x)(\mathbf{l})) \oplus \lambda x\lambda P.\mathbf{onTh}(P(x)(\mathbf{r})); \mathrm{NP\backslash(VP/PP/NP)\backslash VP}\end{matrix}}\backslash\mathrm{E}$$

$$\dfrac{\dfrac{\begin{array}{c}\vdots\quad\vdots\\ \mathsf{to}\circ\mathsf{robin}\circ\mathsf{on}\circ\mathsf{thursday}\circ\\ \mathsf{and}\circ\mathsf{to}\circ\mathsf{leslie}\circ\mathsf{on}\circ\mathsf{friday};\\ [\lambda x\lambda P\lambda z.\mathbf{onFr}(P(x)(\mathbf{l}))(z)]\\ \oplus[\lambda x\lambda P\lambda z.\mathbf{onTh}(P(x)(\mathbf{r}))(z)];\\ \mathrm{NP}\backslash(\mathrm{VP/PP/NP})\backslash\mathrm{VP}\end{array}\qquad\dfrac{\begin{array}{ll}\lambda\sigma.\sigma(\mathsf{the}\circ & \vdots\quad\vdots\\ \mathsf{same}\circ\mathsf{thing}); & \lambda\varphi_1\lambda\varphi_2.\mathsf{terry}\circ\mathsf{said}\circ\varphi_1\circ\varphi_2;\\ \mathbf{same}(\mathbf{thing}); & \lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t});\\ (\mathrm{S|X})|(\mathrm{S|X|NP}) & \mathrm{S|(NP\backslash(VP/PP/NP)\backslash VP)|NP}\end{array}}{\lambda\varphi_2.\mathsf{terry}\circ\mathsf{said}\circ\mathsf{the}\circ\mathsf{same}\circ\mathsf{thing}\circ\varphi_2;\ \mathbf{same}(\mathbf{thing})(\lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t}));\ \mathrm{S|(NP\backslash(VP/PP/NP)\backslash VP)}}\,|\mathrm{E}}{\begin{array}{l}\mathsf{terry}\circ\mathsf{said}\circ\mathsf{the}\circ\mathsf{same}\circ\mathsf{thing}\circ\\ \mathsf{to}\circ\mathsf{robin}\circ\mathsf{on}\circ\mathsf{thursday}\circ\mathsf{and}\circ\mathsf{to}\circ\mathsf{leslie}\circ\mathsf{on}\circ\mathsf{friday};\\ \mathbf{same}(\mathbf{thing})(\lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t}))([\lambda x\lambda P\lambda z.\mathbf{onFr}(P(x)(\mathbf{l}))(z)]\\ \oplus[\lambda x\lambda P\lambda z.\mathbf{onTh}(P(x)(\mathbf{r}))(z)]);\mathrm{S}\end{array}}\,|\mathrm{E}$$

What is crucial in this derivation is the part that derives the NCC involving the strings *to Robin on Thursday* and *to Leslie on Friday*. These strings are analyzed as (nonstandard) constituents via hypothetical reasoning in the same way as other examples above. They are then coordinated with the generalized sum meaning of *and* to form a (generalized) sum of type $e \to (e \to e \to e \to t) \to e \to t$ objects. The rest of the derivation involves creating a doubly-abstracted proposition by abstracting over the positions corresponding to the NP containing *same* and the (higher-order) sum-denoting expression and giving this proposition as an argument to the symmetrical predicate together with its other two arguments, namely, the (higher-order) sum derived above and the noun that provides the descriptive content for the unique entity involved. The translation for the whole sentence is unpacked and simplified in (30):

(30) $\mathbf{same}(\mathbf{thing})(\lambda x\lambda f.f(x)(\mathbf{say})(\mathbf{t}))([\lambda x\lambda P\lambda z.\mathbf{onFr}(P(x)(\mathbf{l}))(z)]\oplus[\lambda x\lambda P\lambda z.\mathbf{onTh}(P(x)(\mathbf{r}))(z)])$

$= \exists y[\mathbf{thing}(y) \wedge \forall R.R <_a [\lambda xLPLz.\mathbf{onFr}(P(x)(\mathbf{l}))(z)]$
$\oplus[\lambda xLP\lambda z.\mathbf{onTh}(P(x)(\mathbf{r}))(z)] \to R(y)(\mathbf{say})(\mathbf{t})]$

$= \exists y[\mathbf{thing}(y) \wedge \lambda x\lambda P\lambda z.[\mathbf{onFr}(P(x)(\mathbf{l}))(z)](y)(\mathbf{say})(\mathbf{t}) \wedge$
$\lambda x\lambda P\lambda z.[\mathbf{onTh}(P(x)(\mathbf{r}))(z)](y)(\mathbf{say})(\mathbf{t})]$

$= \exists y[\mathbf{thing}(y) \wedge \mathbf{onFr}(\mathbf{say}(y)(\mathbf{l}))(\mathbf{t}) \wedge \mathbf{onTh}(\mathbf{say}(y)(\mathbf{r}))(\mathbf{t})]$

This asserts the existence of some unique entity which was said by Terry both to Robin on Thursday and to Leslie on Friday. This correctly corresponds to the internal reading of the sentence where the matters communicated to the two people by Robin on different days are identical to each other.

4.2 UCC and extraction

The interaction between UCC and extraction also receives a straightforward solution in our approach. For the analysis of UCC, we adopt the proposal by Morrill (1994) and Bayer (1996) that involves extending the syntactic type system with the $\vee$ (join) connective. $\vee$ is a two place connective and the complex syntactic category $A \vee B$ intuitively means

that the linguistic expression that is assigned this category belongs to either category A or B. For the join connective, we posit the following two Introduction rules in our system:

(31) a. *Right Join Introduction*

$$\frac{\boldsymbol{a};\boldsymbol{f};A}{\boldsymbol{a};\boldsymbol{f};A\vee B}\vee\mathrm{I}$$

b. *Left Join Introduction*

$$\frac{\boldsymbol{a};\boldsymbol{f};B}{\boldsymbol{a};\boldsymbol{f};A\vee B}\vee\mathrm{I}$$

Intuitively, these rules say that if something is an A (or a B), then we are entitled to conclude a weaker statement that it is $A \vee B$ (i.e. A or B).

The key assumption in the Morrill/Bayer analysis of UCC is the specification of the copula given in (32):

(32) $\mathsf{is};\ \lambda f.f;\ \mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP})$

This says that *is* is looking for either an NP or an AP as its complement to become a VP. With the $\vee$-Introduction rule in (31), the derivation for a sentence in which the copula combines with an NP complement (without UCC) goes as follows:

(33)

$$\dfrac{\mathsf{pat};\mathrm{NP}\qquad\dfrac{\mathsf{is};\mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP})\qquad\dfrac{\mathsf{a}\circ\mathsf{republican};\mathrm{NP}}{\mathsf{a}\circ\mathsf{republican};\mathrm{NP}\vee\mathrm{AP}}\vee\mathrm{I}}{\mathsf{is}\circ\mathsf{a}\circ\mathsf{republican};\mathrm{VP}}/\mathrm{E}}{\mathsf{pat}\circ\mathsf{is}\circ\mathsf{a}\circ\mathsf{republican};\mathrm{S}}\backslash\mathrm{E}$$

The key point here is that, with $\vee$-Introduction, we can assign the category $\mathrm{NP}\vee\mathrm{AP}$ to the string *a Republican* and this satisfies the subcategorization requirement of the copula.

From this, it should already be clear how examples of UCC like *Pat is a Republican and proud of it* are derived. The derivation is given in (34).

(34)

$$\dfrac{\mathsf{pat};\mathrm{NP}\qquad\dfrac{\mathsf{is};\mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP})\qquad\dfrac{\dfrac{\mathsf{a}\circ\mathsf{republican};\mathrm{NP}}{\mathsf{a}\circ\mathsf{republican};\mathrm{NP}\vee\mathrm{AP}}\vee\mathrm{I}\qquad\dfrac{\mathsf{and};(X\backslash X)/X\qquad\dfrac{\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{AP}}{\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{NP}\vee\mathrm{AP}}\vee\mathrm{I}}{\mathsf{and}\circ\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};(\mathrm{NP}\vee\mathrm{AP})\backslash(\mathrm{NP}\vee\mathrm{AP})}/\mathrm{E}}{\mathsf{a}\circ\mathsf{republican}\circ\mathsf{and}\circ\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{NP}\vee\mathrm{AP}}\backslash\mathrm{E}}{\mathsf{is}\circ\mathsf{a}\circ\mathsf{republican}\circ\mathsf{and}\circ\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{VP}}/\mathrm{E}}{\mathsf{pat}\circ\mathsf{is}\circ\mathsf{a}\circ\mathsf{republican}\circ\mathsf{and}\circ\mathsf{proud}\circ\mathsf{of}\circ\mathsf{it};\mathrm{S}}\backslash\mathrm{E}$$

Here, both of the two conjuncts (i.e. the NP *a Republican* and the AP *proud of it*) are derived as NP$\vee$AP via $\vee$-Introduction. Then, with the standard generalized conjunction category for *and*, they are coordinated to form a larger constituent of category NP$\vee$AP. Since this category exactly matches the category that the copula is looking for as its argument, this UCC constituent can be directly combined with the copula via Slash Elimination to complete the derivation. Unlike the ellipsis-based approach of the kind advocated in the LBE

literature, the Morrill/Bayer analysis of UCC in CG treats strings like *a Republican and proud of it* as full-fledged surface constituents without any deletion operation or dummy syntactic head of any kind. As will become clear below, this turns out to be crucial in the analysis of the interactions between UCC and extraction.

For the analysis of extraction, we adopt the proposal by Muskens (2003) that exploits the order-insensitive nature of the non-directional mode of implication (i.e. our vertical slash). In the TLCG literature, the treatment of extraction has long been known as a problematic issue. Essentially, the problem is that modelling extraction by means of forward and backward slashes makes it difficult treat cases of extraction from non-peripheral positions. This is because Slash Introduction can apply for the forward and backward slashes only when (the phonology of) the hypothesis appears at a peripheral position. (An analogous problem is found with CCG, which deals with non-peripheral extraction via order-disrupting, non-harmonic function composition rules.) Various mechanisms have been proposed in the TLCG literature to overcome this problem, but they all involve significant complications in the mapping between syntax and surface morpho-phonology. Muskens's proposal is unique in that it solves this problem by directly representing the phonology of gapped sentences via a higher-order functional phonological term, using a mechanism independently needed in the grammar, namely, λ-binding in phonology. This simplifies the treatment of filler-gap dependency in the CG-based setup considerably.

The core idea of Muskens's (2003) approach to extraction involves analyzing (incomplete) sentences with gaps like *Kim likes* in the topicalization sentence in (35) as a sentence missing some expression somewhere inside, with hypothetical reasoning for the vertical slash, as in the derivation in (36).

(35) Bagels$_i$, Kim likes t_i

(36)

$$\dfrac{\dfrac{\mathsf{bagels};\ \mathbf{b};\ \mathrm{NP} \qquad \dfrac{\lambda\sigma\lambda\varphi.\varphi\circ\sigma(\boldsymbol{\epsilon});\ \lambda f.f;\ (\mathrm{S}|\mathrm{X})|(\mathrm{S}|\mathrm{X}) \qquad \dfrac{\dfrac{\mathsf{kim};\ \mathbf{k};\ \mathrm{NP} \qquad \dfrac{\mathsf{likes};\ \mathbf{like};\ (\mathrm{NP}\backslash\mathrm{S})/\mathrm{NP} \qquad \left[\begin{smallmatrix}\varphi;\\ x;\ \mathrm{NP}\end{smallmatrix}\right]^1}{\mathsf{likes}\circ\varphi;\ \mathbf{like}(x);\ \mathrm{NP}\backslash\mathrm{S}}/\mathrm{E}}{\mathsf{kim}\circ\mathsf{likes}\circ\varphi;\ \mathbf{like}(x)(\mathbf{k});\ \mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi.\mathsf{kim}\circ\mathsf{likes}\circ\varphi;\ \lambda x.\mathbf{like}(x)(\mathbf{k});\ \mathrm{S}|\mathrm{NP}}|\mathrm{I}^1}{\lambda\varphi.\varphi\circ\mathsf{kim}\circ\mathsf{likes};\ \lambda x.\mathbf{like}(x)(\mathbf{k});\ \mathrm{S}|\mathrm{NP}}}{\mathsf{bagels}\circ\mathsf{kim}\circ\mathsf{likes};\ \lambda x.\mathbf{like}(\mathbf{b})(\mathbf{k});\ \mathrm{S}}|\mathrm{E}$$

In (36), an NP is hypothesized in the object position of the transitive verb, and by withdrawing this hypothesis after the whole sentence is built, the position of the gap in the whole string is explicitly represented by the phonology of the hypothesized NP bound by the λ-operator, namely, the variable φ. (Note also that this gapped sentence is assigned the right meaning, that is, the property of being an object that Kim likes, with lambda abstraction of the variable x over the meaning of the whole sentence.) Since hypothetical reasoning for the vertical slash can be carried out regardless of the position of the variable in the surface string (unlike for the directionality-sensitive, forward and backward slashes), this approach can treat filler-gap dependency in a fully general manner wherever the gap appears within

the sentence.

For topicalization, the filler that corresponds to the gap appears in a position immediately to the left of the gapped sentence in the surface string. There is thus a mismatch between the surface form of the sentence and the phonology of the gapped constituent (which takes some string as an argument and embeds it in the original gap site), and this mismatch is mediated by the following phonologically empty topicalization operator:

(37) $\lambda\sigma\lambda\varphi.\varphi \circ \sigma(\boldsymbol{\epsilon});\ \lambda f.f;\ (\mathrm{S}|\mathrm{X})|(\mathrm{S}|\mathrm{X})$

The topicalization operator in (37) does not have any effect on either syntactic category or semantics; it only changes the phonology of the expression it combines with in such a way that, by combining with this topicalization operator, an empty string $\boldsymbol{\epsilon}$ is embedded in the original gap site and the host sentence now concatenates the phonology of its argument (i.e. the filler) immediately to the left of its own phonology.

With this analysis of topicalization, sentences like (38) in which a coordinate structure involving UCC gets topicalized can be analyzed as in (39).

(38) [(Both) poor and a Republican]$_i$ you can't possibly be t_i.

(39)

$$\dfrac{\begin{array}{lll}\vdots\ \vdots & \lambda\sigma\lambda\varphi.\varphi\circ\sigma(\boldsymbol{\epsilon});\ (\mathrm{S}|\mathrm{X})|(\mathrm{S}|\mathrm{X}) & \dfrac{\dfrac{\mathsf{you};\ \mathrm{NP} \qquad \dfrac{\mathsf{can't}\circ\mathsf{be};\ \mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP}) \qquad [\varphi;\mathrm{NP}\vee\mathrm{AP}]^1}{\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{VP}}/\mathrm{E}}{\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi.\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{S}|(\mathrm{NP}\vee\mathrm{AP})}|\mathrm{I}^1\\ \mathsf{both}\circ\mathsf{poor}\circ\mathsf{and}\circ\mathsf{a}\circ\mathsf{republican};\ \mathrm{NP}\vee\mathrm{AP} & & \lambda\varphi.\varphi\circ\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be};\mathrm{S}|(\mathrm{NP}\vee\mathrm{AP})\end{array}}{\mathsf{both}\circ\mathsf{poor}\circ\mathsf{and}\circ\mathsf{a}\circ\mathsf{republican}\circ\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be};\mathrm{S}}|\mathrm{E}\ |\mathrm{E}$$

The key step in this derivation is the hypothetical assumption of an expression of category $\mathrm{NP}\vee\mathrm{AP}$ in the gap position. Via hypothetical reasoning for the vertical slash, a gapped sentence of category $\mathrm{S}|(\mathrm{NP}\vee\mathrm{AP})$ can then be derived, which is missing an expression of category $\mathrm{NP}\vee\mathrm{AP}$, i.e., the complement of the copula. As already shown in the derivation of a simpler UCC sentence above in (34), the string *both poor and a Republican*, which appears in the filler position in (39), can be assigned the category $\mathrm{NP}\vee\mathrm{AP}$. Thus, the gap and the filler match in syntactic category and the two can be combined by means of the topicalization operator in the same way as the previous NP topicalization example in (36).

The (truth-conditional) meaning of a topicalization sentence is obtained simply by substituting the meaning of the filler in the gap position of the host sentence. Thus, in (39), the correct meaning is assigned to the whole sentence in which a conjunction of the two properties is predicated of the subject under the scope of modal and negation.

The analysis of pseudo-cleft is similarly straightforward. Again, the key assumption is the treatment of the gapped sentence with the vertical slash. The gapped sentence is derived in the category $\mathrm{S}|X$, a sentence missing an X somewhere inside. We assign the syntactic category $X|(\mathrm{S}|X)$ to the word *what*, which combines with this gapped sentence and forms the constituent that occupies the precopular position. As illustrated in the following derivation, by assigning this category to *what*, the constituent in the precopular position ends up having the same syntactic category as the gap. Then, with the syntactic category $(X\backslash\mathrm{S})/X$ for the copula, which identifies the categories of its left-hand and right-hand arguments, it follows that the syntactic categories of the gap and the postcopular expression are required to match with one another, capturing the basic syntactic properties of the pseudo-cleft construction. The derivation for the sentence *What Robin wanted was a textbook* is shown in (40):

(40)

$$\dfrac{\dfrac{\lambda\sigma.\mathsf{what}\circ\sigma(\boldsymbol{\epsilon});\ \mathrm{X}|(\mathrm{S}|\mathrm{X}) \qquad \dfrac{\dfrac{\mathsf{robin};\ \mathrm{NP} \qquad \dfrac{\mathsf{wanted};\ (\mathrm{NP}\backslash\mathrm{S})/\mathrm{NP} \qquad [\varphi;\mathrm{NP}]^1}{\mathsf{wanted}\circ\varphi;\mathrm{NP}\backslash\mathrm{S}}/\mathrm{E}}{\mathsf{robin}\circ\mathsf{wanted}\circ\varphi;\mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi.\mathsf{robin}\circ\mathsf{wanted}\circ\varphi;\mathrm{S}|\mathrm{NP}}|\mathrm{I}^1}{\mathsf{what}\circ\mathsf{robin}\circ\mathsf{wanted};\mathrm{NP}}|\mathrm{E} \qquad \dfrac{\mathsf{was};\ (\mathrm{X}\backslash\mathrm{S})/\mathrm{X} \qquad \mathsf{a}\circ\mathsf{textbook};\ \mathrm{NP}}{\mathsf{was}\circ\mathsf{a}\circ\mathsf{textbook};\mathrm{NP}\backslash\mathrm{S}}/\mathrm{E}}{\mathsf{what}\circ\mathsf{robin}\circ\mathsf{wanted}\circ\mathsf{was}\circ\mathsf{a}\circ\mathsf{textbook};\mathrm{S}}|\mathrm{E}$$

As in the analysis of topicalization, the interaction between pseudocleft and UCC is straightforward. By assuming an expression of the complex syntactic category $\mathrm{NP}\vee\mathrm{AP}$ in the gap position, the precopular constituent headed by *what* is derived in the same category as the gap, namely, $\mathrm{NP}\vee\mathrm{AP}$. And then, with the polymorphic syntactic category for the copula, the syntactic category of the precopular and postcopular expressions (which is a UCC of category $\mathrm{NP}\vee\mathrm{AP}$) are identified with each other to complete the derivation. Again, with the assumption that the UCC category denotes a conjunction of two properties, the right semantics is assigned for (11a), where the conjunction of two properties scopes below the negation and the modal.

(41)

$$\dfrac{\dfrac{\lambda\sigma.\mathsf{what}\circ\sigma(\boldsymbol{\epsilon});\ \mathrm{X}|(\mathrm{S}|\mathrm{X}) \qquad \dfrac{\dfrac{\mathsf{you};\ \mathrm{NP} \qquad \dfrac{\mathsf{can't}\circ\mathsf{be};\ \mathrm{VP}/(\mathrm{NP}\vee\mathrm{AP}) \qquad [\varphi;\mathrm{NP}\vee\mathrm{AP}]^1}{\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{VP}}/\mathrm{E}}{\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{S}}\backslash\mathrm{E}}{\lambda\varphi.\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be}\circ\varphi;\mathrm{S}|(\mathrm{NP}\vee\mathrm{AP})}|\mathrm{I}^1}{\mathsf{what}\circ\mathsf{you}\circ\mathsf{can't}\circ\mathsf{be};\mathrm{NP}\vee\mathrm{AP}}|\mathrm{E} \qquad \dfrac{\mathsf{is};\ (\mathrm{X}\backslash\mathrm{S})/\mathrm{X} \qquad \begin{array}{c}\vdots\ \ \vdots\\ \mathsf{intelligentoand}\\ \mathsf{aofundamentalist};\\ \mathrm{NP}\vee\mathrm{AP}\end{array}}{\mathsf{is}\circ\mathsf{intelligent}\circ\mathsf{and}\circ\mathsf{a}\circ\mathsf{fundamentalist};\ (\mathrm{NP}\vee\mathrm{AP})\backslash\mathrm{S}}/\mathrm{E}}{\mathsf{what}\circ\mathsf{youcan't}\circ\mathsf{be}\circ\mathsf{is}\circ\mathsf{intelligent}\circ\mathsf{and}\circ\mathsf{a}\circ\mathsf{fundamentalist};\mathrm{S}}|\mathrm{E}$$

The join connective introduced above in the analysis of UCC has as its dual the meet connective. We will show in Appendix A that the use of this meet connective enables an analysis of examples of UCC with different subcategorization frames such as the following, first noted by Crysmann (2003) and taken to exemplify the superiority of an ellipsis-based analysis of coordination over the direct coordination analysis in CG.

(42) John gave Mary a book and to Peter a record.

4.3 Nominal head coordination

Finally, we analyze the apparent agreement mismatch between the determiner and the verb in nominal head coordination in examples like the following:

(43) That man and woman are arguing again.

Note first that the acceptability of this nominal head coordination pattern partly depends on the semantic/pragmatic properties of the conjoined nominals; as shown in the following examples, combinations of nouns that can naturally be thought of as forming pairs (e.g. *man and woman*, *boy and girl*, *table and chair*) can felicitously appear in this construction, whereas random combinations of nouns (e.g. *man and chair*, *table and boy*) that cannot naturally be construed as forming pairs are generally infelicitous in this construction.

(44) a. This $\left\{\begin{array}{c}\text{man and woman}\\ \text{boy and girl}\\ \text{table and chair}\end{array}\right\}$ are in perfect match.

b.??This $\left\{\begin{array}{c}\text{man and chair}\\ \text{table and boy}\end{array}\right\}$ are in perfect match.

We take this to indicate that the coordinated nominals such as *man and woman* in (43) are a special kind of pair-denoting nominals rather than simply an elliptical version of coordination of full-fledged NPs. (That is, if these examples were derived via ellipsis from coordination of full-fledged NPs, the acceptability contrast in (44) would be puzzling.)

The simplest way to capture this special property of nominal head coordination is to assume that the relevant semantic/pragmatic restriction is encoded in the definition of the covert operator that is responsible for converting the original meanings of such coordinated nominals to the appropriate pair-denoting meanings. With the generalized sum meaning for *and*, 'property sum' meanings of the following form are freely available for coordinated nominals like *man and woman*:

(45) $[\![\text{man and woman}]\!] = \lambda f \lambda g.[f \oplus g](\mathbf{man})(\mathbf{woman}) = \mathbf{man} \oplus \mathbf{woman}$

We posit a following phonologically empty pair-forming operator in (46) that takes such property sums as arguments and returns a property that holds of a pair of individuals just in case the pair in question each satisfy one of the two properties that are parts of the original property sum. The pair-forming operator additionally imposes a semantic/pragmatic restriction such that the original property sum constitutes a 'natural pair' (via the primitive predicate **natural-pair**, which we do not attempt to analyze further here).

(46) $\lambda\varphi.\varphi$; $\lambda P\lambda X.\mathbf{resp}(X)(P) \wedge \mathbf{natural\text{-}pair}(P)$; $X|X$

The pair-forming operator has as its core meaning the definition of the **resp** operator that is essentially identical to the one employed by Gawron & Kehler (2004) for the analysis of 'respective' sentences. The **resp** operator is defined as in (47):

(47) $\mathbf{resp}(X)(P) = 1$ iff
$\exists x, y[\mathbf{atom}(x) \wedge \mathbf{atom}(y) \wedge x \neq y \wedge X = x \oplus y \wedge \exists p, q <_a P[p \neq q \wedge p(x) \wedge q(y)]]$

$\mathbf{resp}(X)(P)$ is true of a sum of individuals X and a sum of properties P just in case there is a bijective relation between the set of individuals that are atomic subparts of X and the set of properties that are atomic subparts of P, such that for each such pair, the individual in question satisfies the property in question.

By applying the pair-forming operator to the property sum meaning of the nominal head coordination *man and woman*, we get the following set of pairs of individuals as output, which is a set of man-woman pairs:

(48) $[\![(46)]\!]([\![\text{man and woman}]\!])$
= $\lambda P\lambda X.[\mathbf{resp}(X)(P) \wedge \mathbf{natural\text{-}pair}(P)](\mathbf{man} \oplus \mathbf{woman})$
= $\lambda X.[\mathbf{resp}(X)(\mathbf{man} \oplus \mathbf{woman}) \wedge \mathbf{natural\text{-}pair}(\mathbf{man} \oplus \mathbf{woman})]$

On this approach, symmetrical modifiers such as *mutually incompatible*, which pose problems for the ellipsis-based analysis, can be treated simply as intersective modifiers that restrict the set of pairs (or groups) of individuals denoted by the head noun. Specifically, the meaning of *mutually incompatible* is given in (49), which takes a set of pairs (or groups) of individuals and imposes on these pairs (or groups) the further condition that they consist of members that are incompatible with each other.

(49) mutually ∘ incompatible; $\lambda P\lambda X.P(X) \wedge \mathbf{incompatible}(X)$; $\mathrm{N/N}$

With these assumptions about the denotations of coordinated nominal heads and symmetrical modifiers, the analysis for (14) goes as in (50):

(50)

$$\dfrac{\text{mutually} \circ \text{incompatible};\ \lambda P\lambda X.P(X) \wedge \mathbf{incmptbl}(X);\mathrm{N/N} \qquad \dfrac{\begin{array}{c}\vdots\ \ \vdots \\ \text{man} \circ \text{and} \circ \text{woman};\ \mathbf{man} \oplus \mathbf{woman};\mathrm{N} \qquad \lambda\varphi_1.\varphi_1;\ \lambda P\lambda X.\mathbf{resp}(X)(P);X|X\end{array}}{\text{man} \circ \text{and} \circ \text{woman};\ \lambda X.\mathbf{resp}(X)(\mathbf{man} \oplus \mathbf{woman});\mathrm{N}}\,|\mathrm{E}}{\text{mutually} \circ \text{incompatible} \circ \text{man} \circ \text{and} \circ \text{woman};\ \lambda X.\mathbf{resp}(X)(\mathbf{man} \oplus \mathbf{woman}) \wedge \mathbf{incmptbl}(X);\mathrm{N}}\,/\mathrm{E}$$

This denotes a set of man-woman pairs such that for each pair, the two individuals constituting the pair are incompatible with one another. The determiner *this* picks up a unique member from this set that is proximal to the speaker.

We take the apparent agreement mismatch between the determiner and the verb to receive a semantic account along the following lines. The singular agreement between the determiner and the coordinated nominal reflects the selectional restriction that the singular determiner imposes on the head noun such that the number of object(s) that satisfy the property denoted by the (coordinated) head noun is one. The plural agreement between the whole NP and the verb, on the other hand, reflects the number of object(s) (in terms of atomic individuals of type e) for which the verbal predicate holds. The man-woman pair that the subject NP denotes is semantically a sum of individuals (just like other plural NPs such as *John and Bill*), and thus triggers plural verb agreement. In short, there is an (apparent) agreement mismatch here since, for pair-denoting nominals, the determiner counts the number of pairs whereas the verb counts the number of members that constitute the pair(s).[6]

To summarize, here again, the right analysis that enables a systematic treatment of a wider range of facts involving symmetrical modifiers is not in terms of ellipsis, but one which directly assigns meanings to such apparently anomalous coordinate structures by means of (slight extensions of) independently motivated mechanisms of grammar such as the generalized sum meaning for *and* and the **resp** operator used in the analysis of 'respectively' sentences. Furthermore, the apparently anomalous agreement pattern, which at first sight appears to motivate an ellipsis-based analysis, receives a fully coherent account by means of an interaction between relevant syntactic and semantic factors.

5 Conclusion

In this paper, we discussed three cases in which non-standard coordination interacts with scopal expressions. The empirical generalization that emerges in these three cases is uniform: the scopal operator that appears outside the coordinate structure in the overt string always takes scope over the whole coordinate structure—in other words, the surface form of the sentence transparently reflects the relevant scopal relation. The null hypothesis in such a situation is that the syntactic constituency relevant for semantic interpretation mirrors this surface constituency between the coordinate structure and the scopal expression.

An ellipsis-based analysis of coordination like the LBE approach in the recent

[6]Heycock & Zamparelli (2005) argue against an analysis of data like (43) which posits an empty pair-forming operator (superficially) similar to our (46), by giving five reasons for rejecting such an analysis. All of their arguments crucially rest on the assumption that the inaudible pair-forming operator has exactly the same syntactic, semantic and morphological properties as the overt word *pair*. But note that such an assumption is dubious given that the distribution of expressions like *this man and woman (are)* is rather restricted (i.e. limited to cases that can be informally described by the notion of 'natural pair', as discussed in the main text) as compared to the overt noun *pair*, which does not come with any such restriction. For this reason, we take it that the facts discussed by Heycock and Zamparelli do not undermine our analysis.

HPSG literature goes wrong for this very reason. Specifically, in this type of approach, if the surface form of the sentence demands an analysis in which the scopal operator is part of the material that undergoes surface ellipsis, the default prediction is that the scopal operator takes scope *inside* each conjunct, but such readings are systematically lacking in all of the three cases considered above. For a similar discrepancy between the underlying combinatoric structure and the actual interpretation observed with generalized quantifiers, Beavers & Sag (2004) propose a mechanism called Optional Quantifier Merger, which specifically does away with the duplication of quantifier meanings from the final interpretation of the sentence on the condition that surface ellipsis takes place—effectively stipulating by fiat the effect that one would automatically get if the surface form of the sentence was directly assigned semantic interpretation without the ellipsis mechanism. Beavers & Sag's (2004) approach covers only the case of generalized quantifiers and it is not clear if it is extendable to other cases like those discussed in the present paper—in particular those involving symmetrical predicates, whose semantics is known to be more complex than that of ordinary generalized quantifiers (Keenan 1992; Barker 2007).

As we have discussed, CG offers a potentially very promising framework for analyzing these complex interactions between coordination and scopal expressions, given its transparent syntax-semantics interface and given its renowned 'direct coordination' analysis of non-standard coordination. However, the standard directional variants of CG is less than optimal for the treatment of scopal expressions due to the fact that the basic mode of implication dealing with syntactic combinatorics is inherently sensitive to word order. Thus, in order to analyze the interactions between scopal expressions and coordination in a fully general manner, we have chosen to extend a directional fragment, which is essentially a labelled deduction (re)formulation of the Lambek calculus, with a mechanism that deals with directionality-insensitive reasoning, incorporating the insight of Oehrle's (1994) term-labelled calculus for quantification. The resultant system recognizes both directional and non-directional modes of implication within a single calculus, and the two types of inference feed into one another freely. As we have shown above, this hybrid architecture of the present framework plays a crucial role in capturing the the empirical interactions between coordination (whose analysis involves inferences with the directional mode of implication) and scope-taking expressions (whose analysis involves the non-directional mode of implication). We thus conclude that the direct coordination analysis of non-standard coordination in CG is truly superior to an alternative that extensively relies on surface ellipsis like the LBE approach in the recent HPSG literature, but that the real empirical payoff of the direct coordination analysis becomes fully apparent only when it is embedded in a framework—like the one we have proposed in this paper—which can deal with complex yet systematic interactions between directional and non-directional modes of inference, each modelling the behaviors of different types of linguistic phenomena in a fully general manner.

References

ABBOTT, BARBARA. 1976. Right node raising as a test for constituenthood. *Linguistic Inquiry* 7.639–642.

BARKER, CHRIS. 2007. Parasitic scope. *Linguistics and Philosophy* 30.407–444.

BAYER, SAMUEL. 1996. The coordination of unlike categories. *Language* 72.579–616.

BEAVERS, JOHN, & IVAN A. SAG. 2004. Coordinate ellipsis and apparent non-constituent coordination. In *The Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Stefan Müller, 48–69, Stanford. CSLI.

BERNARDI, RAFFAELLA. 2002. *Reasoning with Polarity in Categorial Type Logic*. University of Utrecht dissertation. [Available at `http://www.inf.unibz.it/~bernardi/finalthesis.html`].

CARLSON, GREG N. 1987. Same and different: Some consequences for syntax and semantics. *Linguistics and Philosophy* 10.531–565.

CHAVES, RUI PEDRO. 2006. Coordination of unlikes without unlike categories. In *The Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Stefan Müller, 102–122, Stanford. CSLI Publications.

——. 2007. *Coordinate Structures - Constraint-based Syntax-Semantics Processing*. Portugal: University of Lisbon dissertation.

——. 2012. Conjunction, cumulation and respectively readings. *Journal of Linguistics* 48.297–344.

CRYSMANN, BERTHOLD. 2003. An asymmetric theory of peripheral sharing in HPSG: Conjunction reduction and coordination of unlikes. In *Proceedings of Formal Grammar 2003*, ed. by Gerhard Jäger, Paola Monachesi, Gerald Penn, & Shuly Wintner, 47–62. Available at http://cs.haifa.ac.il/~shuly/fg03/.

DE GROOTE, PHILIPPE. 2001. Towards abstract categorial grammars. In *Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference*, 148–155.

DOWTY, DAVID. 1988. Type raising, functional composition, and non-constituent conjunction. In *Categorial Grammars and Natural Language Structures*, ed. by Richard T. Oehrle, Emmon Bach, & Deirdre Wheeler, 153–198. Dordrecht: D. Reidel Publishing Company.

GAWRON, JEAN MARK, & ANDREW KEHLER. 2004. The semantics of respective readings, conjunction, and filler-gap dependencies. *Linguistics and Philosophy* 27.169–207.

GAZDAR, GERALD. 1981. Unbounded dependencies and coordinate structure. *Linguistic Inquiry* 12.155–184.

HEYCOCK, CAROLINE, & ROBERTO ZAMPARELLI. 2005. Friends and colleagues: Plurality, coordination, and the structure of dp. *Natural Language Semantics* 13.201–270.

JACKENDOFF, RAY. 1977. *X-bar Syntax: A Study of Phrase Structure*. Cambridge, MA, USA: MIT Press.

KATHOL, ANDREAS. 1995. *Linearization-Based German Syntax*. Columbus: Ohio State University dissertation.

KEENAN, EDWARD L. 1992. Beyond the Frege boundary. *Linguistics and Philosophy* 15.199–221.

KUBOTA, YUSUKE. 2010. *(In)flexibility of Constituency in Japanese in Multi-Modal Categorial Grammar with Structured Phonology*. The Ohio State University dissertation.

——. to appear. The logic of complex predicates: A deductive synthesis of 'argument sharing' and 'verb raising'. To appear in *Natural Language and Linguistic Theory*.

——, & ROBERT LEVINE. 2012. Gapping as like-category coordination. In *Logical Aspects of Computational Linguistics: 7th International Conference*, ed. by Denis Béchet & Alexander Dikovsky, 135–150. Springer.

——, & ROBERT LEVINE. 2013. Against ellipsis: Arguments for the direct licensing of 'non-canonical' coordinations. MS., University of Tokyo and Ohio State University.

——, & CARL POLLARD. 2010. Phonological interpretation into preordered algebras. In *The Mathematics of Language: 10th and 11th Biennial Conference*, ed. by Christian Ebert, Gerhard Jäger, & Jens Michaelis, 200–209. Springer.

LAMBEK, JOACHIM. 1958. The mathematics of sentence structure. *American Mathematical Monthly* 65.154–170.

LEVINE, ROBERT. 2011. Linerarization and its discontents. In *The Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Stefan Müller, 126–146, Stanford. CSLI Publications.

MOORTGAT, MICHAEL. 1997. Categorial Type Logics. In *Handbook of Logic and Language*, ed. by Johan van Benthem & Alice ter Meulen, 93–177. Amsterdam: Elsevier.

——, & RICHARD T. OEHRLE. 1994. Adjacency, dependence, and order. In *Proceedings of the Ninth Amsterdam Colloquium*, ed. by Paul Dekker & Martin Stokhof, 447–466, Universiteit van Amsterdam. Instituut voor Taal, Logica, en Informatica.

MORRILL, GLYN, & TERESA SOLIAS. 1993. Tuples, discontinuity, and gapping in categorial grammar. In *Proceedings of the Sixth Conference of the European Chapter of the Association for Computational Linguistics*, 287–297, Morristown, NJ. Association for Computational Linguistics.

MORRILL, GLYN V. 1994. *Type Logical Grammar: Categorial Logic of Signs*. Dordrecht: Kluwer Academic Publishers.

MUSKENS, REINHARD. 2001. Categorial Grammar and Lexical-Functional Grammar. In *The Proceedings of the LFG '01 Conference*, ed. by Miriam Butt & Tracy Holloway King, University of Hong Kong.

——. 2003. Language, lambdas, and logic. In *Resource Sensitivity in Binding and Anaphora*, ed. by Geert-Jan Kruijff & Richard Oehrle, Studies in Linguistics and Philosophy, 23–54. Kluwer.

——. 2007. Separating syntax and combinatorics in categorial grammar. *Research on Language and Computation* 5.267–285.

OEHRLE, RICHARD T. 1994. Term-labeled categorial type systems. *Linguistics and Philosophy* 17.633–678.

PARTEE, BARBARA, & MATS ROOTH. 1983. Generalized quantifiers and type ambiguity. In *Meaning, Use, and Interpretation of Language*, ed. by Rainer Bäuerle, Christoph Schwarze, & Arnim von Stechow, 361–383. Berlin: Walter de Gruyter.

POLLARD, CARL. 2009. Parasitic scope in categorial grammar with φ-labelling. Presentation at the Synners meeting, May 27, 2009.

——. 2011. Proof theoretic background for linear grammar. MS., Ohio State University.

——, & E. ALLYN SMITH. to appear. A unified analysis of *the same*, phrasal comparatives and superlatives. In *Proceedings of SALT 2012*, volume ??, ??–??

REAPE, MIKE. 1996. Getting things in order. In *Discontinuous Constituency*, ed. by Harry Bunt & Arthur van Horck, volume 6 of *Natural Language Processing*, 209–253. Berlin, Germany and New York, NY, USA: Mouton de Gruyter. Published version of a Ms. from 1990.

SAG, IVAN, & RUI CHAVES. 2008. Left- and right-periphery ellipsis in coordinate and non-coordinate structures. MS., Stanford University and University at Buffalo, The State University of New York.

STEEDMAN, MARK. 1985. Dependency and coordination in the grammar of Dutch and English. *Language* 61.523–568.

YATABE, SHÛICHI. 2001. The syntax and semantics of left-node raising in Japanese. In *Proceedings of the 7th International Conference on Head-Driven Phrase Structure Grammar*, ed. by Dan Flickinger & Andreas Kathol, 325–344, Stanford. CSLI. http://cslipublications.stanford.edu/HPSG/.

YATABE, SHÛICHI. 2012. Comparison of the ellipsis-based theory of non-constituent coordination with its alternatives. In *Proceedings of the 19th International Conference on Head-Driven Phrase Structure Grammar, Chungnam National University Daejeon*, ed. by Stefan Müller, 453–473.

ZAENEN, ANNIE, & LAURI KARTTUNEN. 1984. Morphological non-distinctiveness and coordination. In *Proceedings of the First Eastern States Conference on Linguistics*, ed. by Gloria Alvarez, Belinda Brodie, & Terry McCoy, 309–320.

A Coordination of unlikes with different subcategorization frames

In this appendix, we show that, by adopting the 'semantically potent' variant of the meet connective (in Bayer's (1996) terminology), examples such as (51) receives a straightforward analysis in the direct coordination analysis in CG, thereby refuting the claim occasionally raised in the literature by proponents of the LBE approach coordination that such examples undermine the CG analysis of NCC.

(51) John gave Mary a book and to Peter a record.

We assume that (51) is a variant of (52) which has undergone a surface-oriented reordering operation. For expository convenience, we provide the derivation for (52), and gloss over the details of the reordering operation.

(52) ?John gave Mary a book and a record to Peter.

The semantically potent variant of the meet connective assigns pairs of meanings as the denotations of the linguistic expressions that are assigned such categories. Thus, on this analysis, the different subcategorization frames of *give* can be compiled into one lexical entry in the following form:

(53) gave; $\langle \lambda x \lambda y \lambda z.\mathbf{give}(x)(y)(z), \lambda x \lambda y \lambda z.\mathbf{give}(y)(x)(z) \rangle$; (VP/PP/NP)∧(VP/NP/NP)

Note that, corresponding to the two subcategorization frames enocded in the syntactic category $\mathrm{VP/PP/NP}$ and $\mathrm{VP/NP/NP}$, we have two distinct semantic translations involving the same constant **give** but which take the first two arguments in different orders.

In actual derivations, one of these subcategorization frames is chosen via Meet Elimination:

(54)
$$\dfrac{\mathsf{john};\mathbf{j};\mathrm{NP} \quad \dfrac{\dfrac{\dfrac{\mathsf{gave};\ \langle \lambda x \lambda y \lambda z.\mathbf{give}(x)(y)(z), \lambda x \lambda y \lambda z.\mathbf{give}(y)(x)(z) \rangle;\ (\mathrm{VP/PP/NP}) \wedge (\mathrm{VP/NP/NP})}{\mathsf{gave};\ \lambda x \lambda y \lambda z.\mathbf{give}(y)(x)(z);\ \mathrm{VP/NP/NP}}\wedge\mathrm{E} \quad \mathsf{mary};\mathbf{m};\mathrm{NP}}{\mathsf{gave} \circ \mathsf{mary};\ \lambda y \lambda z.\mathbf{give}(y)(\mathbf{m})(z);\ \mathrm{VP/NP}}/\mathrm{E} \quad \mathsf{the} \circ \mathsf{book};\mathbf{b};\mathrm{NP}}{\mathsf{gave} \circ \mathsf{mary} \circ \mathsf{the} \circ \mathsf{book};\ \lambda z.\mathbf{give}(\mathbf{b})(\mathbf{m})(z);\ \mathrm{VP}}/\mathrm{E}}{\mathsf{john} \circ \mathsf{gave} \circ \mathsf{mary} \circ \mathsf{the} \circ \mathsf{book};\ \mathbf{give}(\mathbf{b})(\mathbf{m})(\mathbf{j});\ \mathrm{S}}\backslash\mathrm{E}$$

With these assumptions, the derivation for (52) is now straightforward. It just involves an interaction of the usual hypothetical reasoning analysis of NCC and the meet elimination analysis of the 'disambiguation' of two subcategorization frames assigned to *give* in (53). (Here, DTV abbreviates $\mathrm{VP/NP/NP}$ and PDTV abbreviates $\mathrm{VP/PP/NP}$; π_1 and π_2 are the first and second projection functions.)

(55)

$$\dfrac{\dfrac{\dfrac{\dfrac{[\varphi; f; \mathrm{PDTV} \wedge \mathrm{DTV}]^1}{\varphi; \pi_2(f); \mathrm{DTV}}\wedge\mathrm{E} \quad \mathsf{mary}; \mathbf{m}; \mathrm{NP}}{\varphi \circ \mathsf{mary}; \pi_2(f)(\mathbf{m}); \mathrm{VP/NP}}/\mathrm{E} \quad \mathsf{the} \circ \mathsf{book}; \mathbf{b}; \mathrm{NP}}{\varphi \circ \mathsf{mary} \circ \mathsf{the} \circ \mathsf{book}; \pi_2(f)(\mathbf{m})(\mathbf{b}); \mathrm{VP}}/\mathrm{E}}{\begin{array}{l}\mathsf{mary} \circ \mathsf{the} \circ \mathsf{book};\\ \lambda f.\pi_2(f)(\mathbf{m})(\mathbf{b});\\ (\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP}\end{array}}|\mathrm{I}^1 \quad \dfrac{\begin{array}{l}\mathsf{and};\\ \lambda g\lambda h.g \sqcap h;\\ (X\backslash X)/X\end{array} \quad \begin{array}{c}\vdots\ \ \vdots\\ \begin{array}{l}\mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda f.\pi_1(f)(\mathbf{r})(\mathbf{p});\\ (\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP}\end{array}\end{array}}{\begin{array}{l}\mathsf{and} \circ \mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda h.[\lambda f.\pi_1(f)(\mathbf{r})(\mathbf{p})] \sqcap h;\\ ((\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP})\backslash((\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP})\end{array}}/\mathrm{E}}{\begin{array}{l}\mathsf{mary} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda f.\pi_1(f)(\mathbf{r})(\mathbf{p}) \wedge \pi_2(f)(\mathbf{m})(\mathbf{b});\\ (\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP}\end{array}}\backslash\mathrm{E}$$

$$\dfrac{\begin{array}{l}\mathsf{gave};\\ \langle\lambda x\lambda y\lambda z.\mathbf{give}(x)(y)(z), \lambda x\lambda y\lambda z.\mathbf{give}(y)(x)(z)\rangle;\\ \mathrm{PDTV} \wedge \mathrm{DTV}\end{array} \quad \begin{array}{c}\vdots\ \ \vdots\\ \begin{array}{l}\mathsf{mary} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda f.\pi_1(f)(\mathbf{r})(\mathbf{p}) \wedge \pi_2(f)(\mathbf{m})(\mathbf{b});\\ (\mathrm{PDTV} \wedge \mathrm{DTV})\backslash\mathrm{VP}\end{array}\end{array}}{\begin{array}{l}\mathsf{gave} \circ \mathsf{mary} \circ \mathsf{the} \circ \mathsf{book} \circ \mathsf{and} \circ \mathsf{the} \circ \mathsf{record} \circ \mathsf{to} \circ \mathsf{peter};\\ \lambda x.\mathbf{give}(\mathbf{r})(\mathbf{p})(x) \wedge \mathbf{give}(\mathbf{b})(\mathbf{m})(x);\\ \mathrm{VP}\end{array}}\backslash\mathrm{E}$$

Now, we should recall that, in defining the semantics of the meet connective, Bayer (1996) (section 7.1) opts for an impoverished, semantically nonpotent definition. The alleged motivation for this choice, according to Bayer (1996), comes from the fact that examples like the following (constituting a violation of Zaenen & Karttunen's (1984) well-known Anti-Pun Ordinance) can be derived once we admit the semantically potent variant of the meet connective and assign to *can* a syntactic category $(\mathrm{VP/NP}) \wedge (\mathrm{VP/VP[BASE]})$ with the corresponding semantic interpretation which pairs the main verb meaning and the auxiliary meaning as one entry.

(56) *I can tuna for a living and get a job if I want.

We take it that Bayer's (1996) argument here is somewhat misguided. In particular, Bayer (1996) seems to overlook the point that the availability of the semantically potent variant of the meet connective in the theory does not necessitate an analysis of the ambiguity of *can* in terms of it. The auxiliary *can* and the main verb *can* can just be entered in the lexicon as two separate entries, and then the overgeneration of (56) does not arise.

This of course begs the question of how to restrict the use of the semantically potent variant of the meet connective. In order to provide a complete answer to this question, we need to study examples of the sort exemplified by (52) more closely, but a conceptually plausible hypothesis which gives us a starting point is readily available: what distinguishes the ungrammatical cases like (56) and the grammatical cases like (52) seems to be that,

while the two uses of the same phonological form are totally unrelated in the former, the two subcategorization frames of *give* in the latter are clearly related semantically. By assuming that a semantically potent variant of the meet connective is invoked only in cases like the latter, we have a principled explanation for the Anti-Pun Ordinance while at the same time recognizing semantically potent meet. In other words, Anti-Pun Ordinance is a reflection of some substantive generalization governing the lexicon (whose exact nature we still don't understand), and not a consequence of a formal property of the underlying logic.

PERCEIVED FOREIGN ACCENT IN THREE VARIETIES OF NON-NATIVE ENGLISH*

Elizabeth A. McCullough
Ohio State University

Abstract

What aspects of the speech signal cause listeners to perceive a foreign accent? While many studies have explored this question for a single variety of non-native speech, few have simultaneously considered non-native speech from multiple native language backgrounds. In this perception study, American English-speaking listeners rated stop-vowel sequences extracted from English words produced by L1 American English, L1 Hindi, L1 Korean, and L1 Mandarin talkers on a continuous scale of degree of foreign accent. Stepwise linear regression models revealed that VOT, vowel quality, f0, and vowel duration contributed significantly to the ratings. Additionally, listeners rated productions by all varieties of non-native talkers as sounding foreign-accented to some degree, with those by L1 Hindi talkers as most foreign-accented, and those by L1 Mandarin talkers as more foreign-accented than those by L1 Korean talkers. The results suggest that several acoustic properties contribute substantially to the perception of foreign accent, at least for stop-vowel sequences, and that some varieties of non-native English sound more accented than others.

*I thank Mary Beckman, Cynthia Clopper, and Jeff Holliday for valuable input as this project developed, and audiences at the 161st meeting of the ASA and LabPhon13 for helpful comments on previous versions of this work.

Elizabeth A. McCullough. Perceived foreign accent in three varieties of non-native English. *OSUWPL* Volume 60, Spring 2013, 51-66.

1 Introduction

Native listeners of a language can recognize when someone is speaking that language with a foreign accent, even on the basis of very short samples of speech (Flege, 1984). However, they need not to do so accurately to motivate the study of foreign accent perception. Even when listeners misidentify a native talker as being foreign, or an L2 talker as being native, they are basing these judgments on some principled idea of what constitutes "foreign accent." Further, these judgments have repercussions, as sounding accented can have negative social consequences for the talker (Gluszek & Dovidio, 2010), and a listener viewing a talker as "other" may be biased against understanding the talker's speech (Rubin, 1992). The goals of the present investigation are to identify which acoustic properties contribute to the perception of foreign accent by native listeners of American English, and to explore such listeners' implicit views about the relative degrees of foreign accent in different varieties of non-native speech.

In many studies of perceived foreign accent, native listeners quantify the degree of foreign accent they hear in each production, and relationships between these ratings and talker-specific characteristics are examined. For instance, Oyama (1976) found a clear relationship between perceived foreign accentedness and age of arrival of Italian immigrants to the United States: the earlier an individual had immigrated, the weaker a foreign accent he was later judged to have. Flege, Munro, and MacKay (1995) conducted a similar investigation of Italian-accented English, and found that age of learning accounted for 59% of the variance in perceived foreign accent ratings, with earlier learners sounding more native. Such studies, however, do not address the question of which characteristics directly influence native listeners in their assignments of perceived foreign accent ratings. Listeners have no knowledge of an individual talker's language history, and must be attending to properties of the acoustic signal.

What acoustic properties might contribute to the perception of foreign accent? Traditional accounts of L2 acquisition (e.g., Lado, 1957) refer to cross-language phonological differences and the role of L1 "interference" in production of the L2. This suggests that it might be possible to measure and compare acoustic properties that are likely to differ between a talker's L1 and L2, and correlate the degree of perceived foreign accent with the degree of difference on these specific measures.

While this approach is not new, investigations of the signal have generally focused on single acoustic properties. For instance, VOT has been found to contribute to the perception of foreign accent for voiceless stops in L2 Spanish productions by L1 English speakers (Gonzalez-Bueno, 1997), and in L2 English productions by L1 Brazilian Portuguese speakers (Major, 1987) and L1 Japanese speakers (Riney & Takagi, 1999). Pronunciations of liquids and vowels seemed to influence the perception of foreign accent in L2 English productions by L1 Japanese speakers in Riney, Takagi, and Inutsuka (2005), although this was determined by a phonetically trained listener's auditory analysis rather than any acoustic measurements. Munro and Derwing (2001) found that speaking rate influenced the perception of foreign accent in the speech of L2 English talkers from 12 different L1s.

In addition, such investigations, with the notable exception of Munro and Derwing (2001), have nearly exclusively focused on single varieties of non-native speech. Thus, it is not clear that these studies investigated "foreign accent" generally as opposed to more

specific scales such as "Brazilian Portuguese accent" or "Japanese accent." If the task demanded only comparisons between native talkers and L2 talkers from a single L1 background, listeners might have implemented the more specific scale regardless of the instructions given.

To focus on "foreign accent" in general as opposed to some particular accent, the present study, like Munro and Derwing (2001), uses L2 English productions by talkers of multiple L1s. In addition, multiple acoustic properties are measured and evaluated in relation to listeners' ratings of the degree of foreign accent in each of these productions. The analysis below will identify acoustic properties that seem to influence foreign accent perception, and determine whether some L1 backgrounds give rise to a stronger percept of foreign accent in L2 English than do others.

2 Methods

2.1 Acoustic stimuli

The acoustic stimuli in this experiment were extracted from recordings in the Buckeye GTA Corpus (Hardman, 2010). This corpus includes productions of 64 of the Bamford-Kowal-Bench sentences revised for American English (Bamford & Wilson, 1979) by 24 L1 American English talkers, 19 L1 Hindi talkers, 20 L1 Korean talkers, and 20 L1 Mandarin talkers. All non-native talkers were of a reasonably high English proficiency level, in that they were certified as a Graduate Teaching Associate at The Ohio State University (by a score of 230/300 on the SPEAK test or an "unconditional pass" on the university's Mock Teaching Test) or had scored at least 26/30 on the speaking section of the TOEFL iBT. For this study, 4 female talkers from each of the 4 L1 groups were used, for a total of 16 talkers. Including 4 talkers from each L1 ensured that there was variation within each L1 group, such that any effects of L1 would not be confounded with individually idiosyncratic pronunciations. Although English is commonly spoken in India, none of the 4 L1 Hindi talkers chosen identified English as a native language.

While many previous studies have used sentences as audio stimuli in perceived foreign accent rating tasks, an exhaustive acoustic investigation of sentences would be a daunting undertaking. Units as large as sentences have a relatively large number of possible segmental and suprasegmental cues, many of which might be expected to influence ratings of perceived foreign accent. In order to limit the number of potentially relevant acoustic cues, stop-vowel sequences were chosen as the acoustic stimuli for the present study. Stop-vowel sequences are even smaller than the words rated in Gonzalez-Bueno (1997) and Major (1987), and as such have a comparatively small number of possible cues to explore, although listeners are still able to perceive foreign accent in units this size (Flege, 1984).

Sequences containing stops are particularly interesting given the language backgrounds included in the Buckeye GTA Corpus. Stops in some of these languages differ from those in American English phonologically: at each place of articulation, Korean distinguishes 3 stops (Lee, 1999) and Hindi, 4 (Ohala, 1999), while American English and Mandarin have only 2 (Ladefoged, 1999; Lee & Zee, 2003). Additionally, the stop contrasts in Hindi, Korean, and Mandarin all differ acoustically from the American English contrast. Such differences might be reflected to some degree in the English productions of these non-native talkers.

The stop-vowel productions used as stimuli were extracted from utterance-medial, word-initial contexts in the Buckeye GTA Corpus recordings. Three stop-vowel sequences were chosen for each of the 6 American English stops, for a total of 18 stop-vowel sequences. The limited number of word types in the corpus did not allow for control of the vowel, nor for the exclusion of stop-vowel sequences that were themselves lexical items (/pi/ and /tu/). All 18 stop-vowel sequences and the words they were extracted from are included in Table 1 in the Appendix.

Each of the 16 talkers produced each of the 18 stop-vowel sequences, for a total of 288 audio stimuli. The mean intensity of all audio stimuli was normalized. No fillers were used.

2.2 Participants

28 monolingual American English-speaking listeners participated in the rating study for linguistics course credit.

2.3 Task

The audio stimuli were presented through headphones at individual computer stations. Listeners were asked to rate the degree of foreign accent in each audio stimulus by sliding a bar along a continuous Visual Analog Scale (VAS). The endpoints of the VAS were labeled "no foreign accent" and "strong foreign accent," as shown in Figure 1. Listeners were encouraged to use the scale continuously rather than categorically. VAS was chosen because listeners have fairly sensitive responses to foreign accent rating tasks, as confirmed by the continuous rating scale used by Flege et al. (1995), which might not be sufficiently captured by a Likert scale with a fixed number of intervals. In addition, a continuous rating scale allowed for the possibility of better correlation with the continuous acoustic cues discussed below.

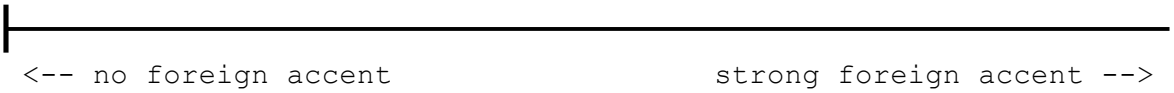

Figure 1. VAS rating line

Just before each stimulus played, the word it had been extracted from was displayed orthographically on the computer screen so that listeners had some perceptual target for the upcoming production. The audio stimuli were blocked by stop-vowel sequence; that is, all 16 productions of the same stop-vowel sequence were played consecutively. Listeners were permitted to take breaks between each block. Block order was randomized, as was stimulus order within each block. A practice block with an additional stop-vowel sequence not used as a test stimulus (/be□/ from baby) was administered before the test blocks so that listeners could experience the types of variation in the stimuli and practice using the rating scale.

2.4 Acoustic properties

Four acoustic properties were considered in the analysis: VOT, vowel quality, vowel duration, and f0 (fundamental frequency). VOT and vowel quality (midpoint F1 and F2)

were chosen as fairly standard measures that have been previously shown to be related to perceived foreign accent ratings (Gonzalez-Bueno, 1997; Major, 1987; Riney & Takagi, 1999; Riney et al., 2005). Vowel duration was included in response to remarks from listeners that some stimuli were too short to rate properly. Similarly, f0 was included in response to remarks from listeners that differences in pitch made the task difficult.

Speaking rate, although found to play a role in foreign accent perception by Munro and Derwing (2001), was not investigated here. Because a temporal measure of each of the two segments in each stop-vowel stimulus was already included, speaking rate would have been largely redundant.

2.5 Analysis

For the production measures, VOT and vowel intervals were marked by the author in Praat (Boersma & Weenink, 2012), and their durations were extracted automatically. F1 and F2 values at the vowel midpoint and mean f0 values over the vowel were extracted automatically using a Praat script that displayed a spectrogram with overlaid formant and pitch tracks, such that the author could check token-by-token for tracking errors. In the rare case of such an error, the author re-measured the token in question with modified settings.

For the perception measure, each listener rating was recorded by the experiment presentation software as a numerical value representing the position on the rating line. In total, 8064 ratings were assigned (16 talkers x 18 stop-vowel sequences x 28 raters).

3 Predictions

While variation in both acoustic measurements and perceived foreign accent ratings can of course be expected in productions by different talkers from the same L1 background and even in different productions by the same talker, this study focuses mainly on variation across L1 backgrounds. Actual language background is expected to align to some degree with perceived foreign accent in a talker's speech, in that productions from native talkers are likely to be judged as exhibiting little to no foreign accent, and productions by non-native talkers are likely to be judged as exhibiting some degree of foreign accent. In this section, the effect of L1 background on each acoustic property is examined for the stimuli used in the present study, and subsequent predictions are made. If there are differences between native and non-native productions for a particular acoustic property, then listeners might use that property to cue foreign accent. Additionally, if a particular variety of non-native speech differs from native speech on multiple acoustic properties, then it might be judged as strongly foreign-accented. As explained further in Section 4.1, stimuli with voiced versus voiceless stop targets will be analyzed separately to facilitate comparison with previous studies; thus, these acoustic measurements are also presented separately for the two groups of stimuli.

3.1 Differences by L1 background

Figure 2 shows VOT values for the productions by talkers of each variety of English. Separate one-way ANOVAs with talkers as subjects and L1 as a between-subjects factor revealed significant effects of L1 on VOT for voiced ($F(3,12) = 6.50$; $p < 0.01$) and voiceless ($F(3,12) = 12.80$; $p < 0.001$) stop targets. Separate Tukey post-hoc tests for

voiced and voiceless stop targets showed that these effects were driven by the differences between L1 Hindi talkers' values and all others ($p < 0.05$); no other pairwise comparisons were significant. Indeed, the most striking details of Figure 2 are the VOT ranges for L1 Hindi talkers' productions. Unlike the voiced stop productions by talkers from other L1 backgrounds, which exhibited short lag voicing, a substantial portion (56%) of the L1 Hindi talkers' voiced stop productions were actually prevoiced. Additionally, L1 Hindi talkers' voiceless stop productions had much shorter VOT values than those for talkers of all other L1s.

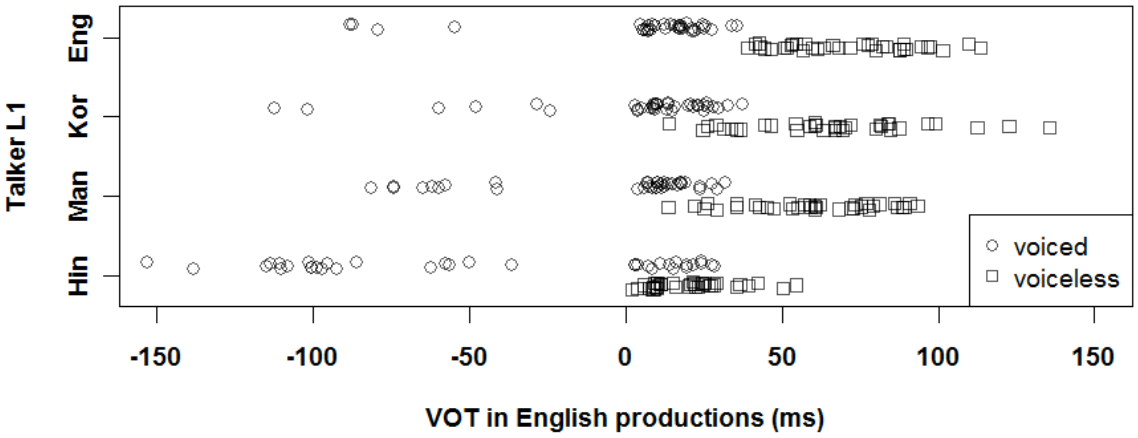

Figure 2. VOT values by target voicing and talker L1

The use of multiple varieties of non-native English in this study presents a problem for linear regression using raw values: productions by native talkers can be expected to have values near one extreme of the range for perceived foreign accent ratings ("no foreign accent"), but the same assumption cannot necessarily be made about the distribution of their acoustic measurements. That is, it would be possible for productions by talkers from one non-native background to have values lower than those observed in productions by native talkers, while productions by talkers from a different non-native background might have values higher than those observed in productions by native talkers. This problem is avoided here by using values that represent the difference between a production and the comparable native productions on some acoustic property. Specifically, all acoustic cues considered below are parameterized as the absolute difference from the American English mean, calculated separately for each of the 18 stop-vowel sequences. For instance, rather than examining the raw VOT value for some L1 Korean talker's production of /pi/ from the word people, the value considered is the absolute difference between the VOT value of this production of /pi/ and the mean VOT of the 4 L1 American English talkers' productions of /pi/. Difference values for productions by L1 American English talkers are also calculated, and should generally be small. Figure 3 shows VOT differences for the productions by each talker. Talkers within each L1 group are presented in a consistent order across all plots.

Separate one-way ANOVAs with talkers as subjects and L1 as a between-subjects factor revealed significant effects of L1 on VOT difference for voiced ($F(3,12) = 7.45$; $p < 0.01$) and voiceless ($F(3,12) = 38.38$; $p < 0.001$) stop targets. Again, this result was due to the values for productions by L1 Hindi talkers being different from those for productions by talkers from all other L1 backgrounds, as shown by separate Tukey post-hoc tests for voiced and voiceless stop targets ($p < 0.05$). In the case of VOT, the effect

of L1 background is the same regardless of whether raw values or values that denote the difference from native production means are used.

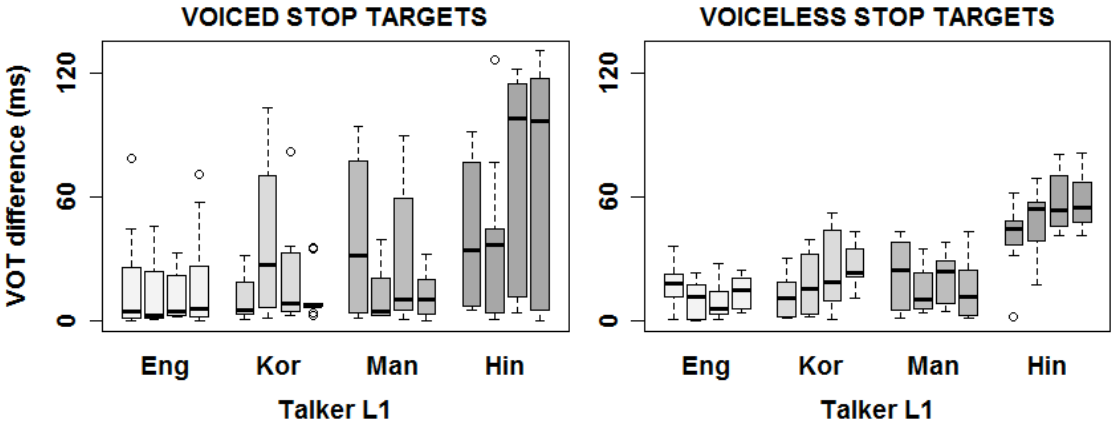

Figure 3. VOT differences by target voicing, talker L1, and talker identity

Vowel quality is defined in terms of two dimensions, F1 and F2. To reduce this to a single dimension, vowel quality difference is quantified as the Euclidean distance in F1/F2 space from the position of a single production to the mean position for the productions of that stop-vowel sequence by the L1 American English talkers. Figure 4 shows vowel quality differences for the productions by each talker. A one-way ANOVA with talkers as subjects and L1 as a between-subjects factor showed a significant effect of L1 on vowel quality difference for stimuli with voiceless stop targets ($F(3,12) = 8.04$; $p < 0.01$); the effect was not significant for stimuli with voiced stop targets. A Tukey post-hoc test for stimuli with voiceless stop targets revealed that vowel quality difference in productions by L1 Korean and L1 Mandarin talkers was different from that in productions by L1 American English talkers ($p < 0.05$); difference values were greater for productions by the non-native talkers. The comparison between difference values for L1 Hindi talkers' productions and L1 American English talkers' productions, in the same direction, approached significance ($p = 0.05$).

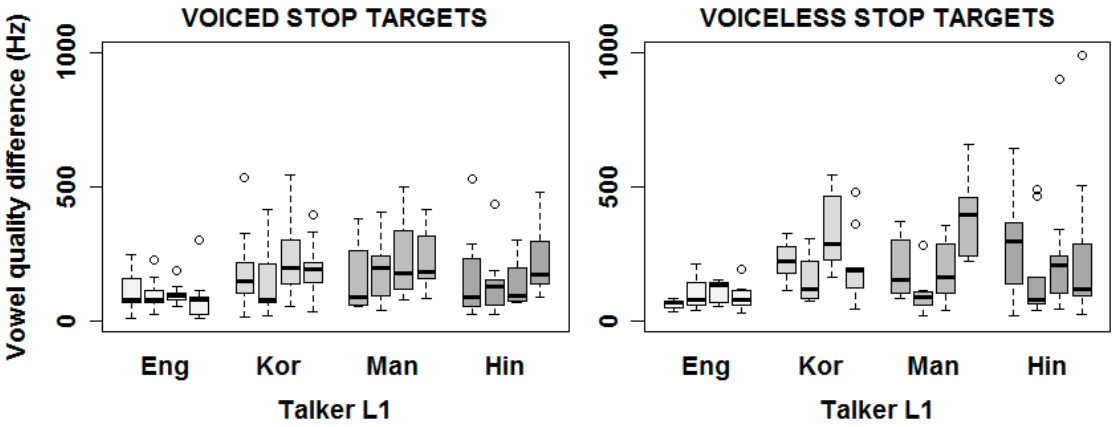

Figure 4. Vowel quality differences by target voicing, talker L1, and talker identity

Figure 5 shows vowel duration differences for the productions by each talker. A one-way ANOVA with talkers as subjects and L1 as a between-subjects factor revealed a

significant effect of L1 on vowel duration difference for stimuli with voiced stop targets ($F(3,12) = 10.12$; $p < 0.01$); the effect was not significant for stimuli with voiceless stop targets. A Tukey post-hoc test for stimuli with voiced stop targets showed that vowel duration difference in productions by L1 Korean and L1 Mandarin talkers was different from that in productions by L1 American English talkers ($p < 0.05$); again, difference values were greater for productions by the non-native talkers.

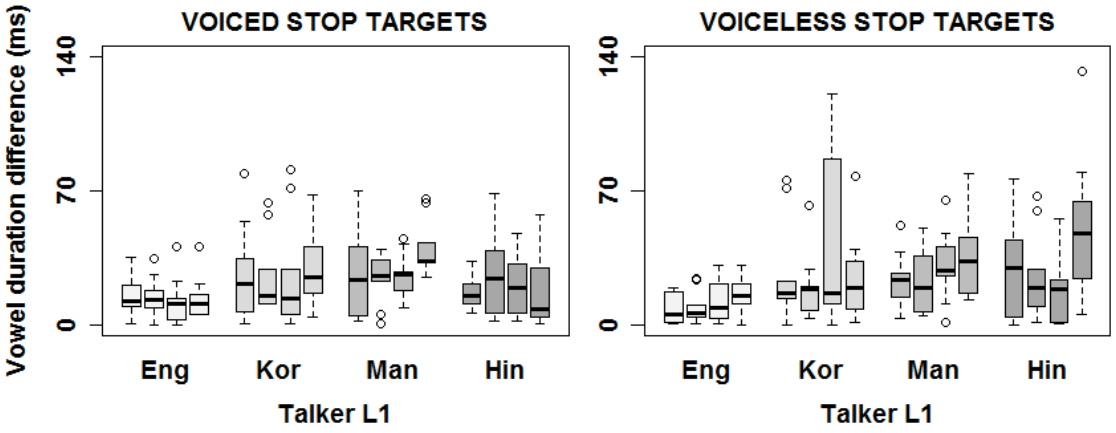

Figure 5. Vowel duration differences by target voicing, talker L1, and talker identity

Figure 6 shows f0 differences for the productions by each talker. Separate one-way ANOVAs with talkers as subjects and L1 as a between-subjects factor did not reveal a significant effect of L1 on f0 difference for either set of stimuli.

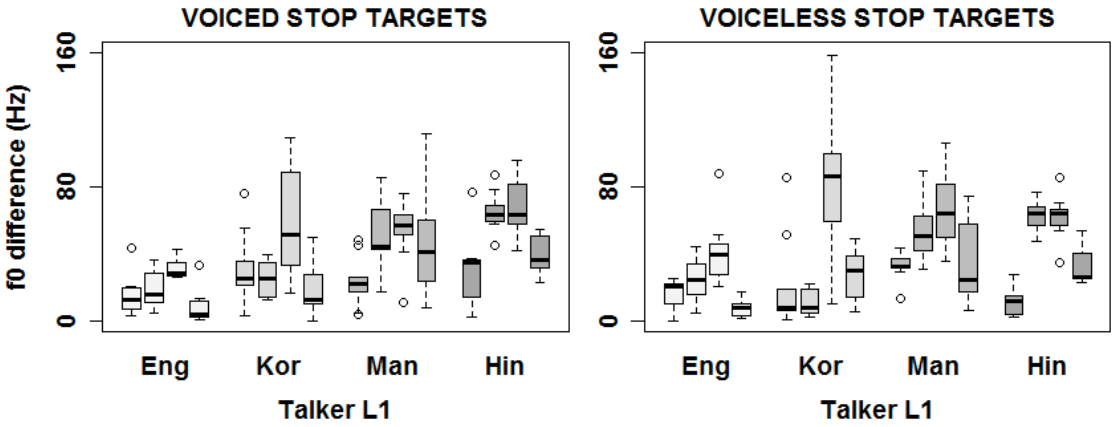

Figure 6. f0 differences by target voicing, talker L1, and talker identity

3.2 Acoustic cues

As differences were observed between native and non-native talkers in VOT, vowel quality, and vowel duration, these acoustic properties might play a role in foreign accent perception. In support of this prediction, the relationship of VOT to foreign accent perception has been demonstrated previously by acoustic measurements (Gonzalez-Bueno, 1997; Major, 1987; Riney & Takagi, 1999), and vowel quality's role has been suggested previously by auditory analysis (Riney et al., 2005). In contrast, f0 did not

reliably differ by L1 background and may be unlikely to influence the perception of foreign accent.

3.3 Overall rating patterns

Productions by L1 Hindi talkers differed from native talkers' productions in VOT and marginally in vowel quality, while productions by L1 Korean and L1 Mandarin talkers differed from native talkers' productions in vowel quality and vowel duration. Due to these acoustic differences, it is expected that productions from all non-native backgrounds should sound accented to some degree. As VOT has repeatedly been shown to correlate with perceived foreign accent ratings, and as productions from L1 Hindi talkers are the only ones that came close to showing effects for properties of both consonants and vowels, it is possible that productions from L1 Hindi talkers might sound more foreign-accented than productions by other non-native talkers. This would align well with feedback from listeners, who often commented that the "Indian accent" was the most obvious.

4 Results

In this section, the possible acoustic cues to foreign accent perception are evaluated, and then the relative strength of foreign accent in different varieties of non-native English is explored.

4.1 Acoustic cues

The regression models detailed in this section test the relationship between perceived foreign accent ratings and acoustic cues directly, and the stepwise models evaluate the contributions of all measured acoustic cues simultaneously. A significant positive relationship between any acoustic cue and perceived foreign accent ratings means that as values of that acoustic cue deviate more from the L1 American English talkers' mean, listeners perceive a higher degree of foreign accent.

The dependent variable is the mean perceived foreign accent rating for each of the 288 stimuli. Thus, while the predictions above were based on differences in acoustic properties across L1 groups, this portion of the analysis should also account for differences in ratings among talkers in the same L1 group, as well as differences in ratings among the 18 productions by a single talker. To the extent that productions varied in some significant acoustic cue, in terms of deviation from the native mean, the ratings for those productions should be different, regardless of the identity of the talker.

Previous studies have found clear relationships between VOT and ratings of perceived foreign accent (Gonzalez-Bueno, 1997; Major, 1987; Riney & Takagi, 1999). Notably, however, these studies only explored this relationship with voiceless stop targets. Thus, the results below are presented separately for voiced and voiceless stop targets, to make comparison with previous findings possible.

A stepwise linear regression model of the data for stimuli with voiced stop targets only identified vowel quality, VOT, f0, and vowel duration, in that order, as significant cues. Four linear regression models were built, such that the first included only the most significant cue, the second included the two most significant cues, etc. Full statistical

details are given in Table 2 in the Appendix. For the present discussion, it is important to highlight that the model with vowel quality accounted for 18% of the variance in the perceived foreign accent ratings. With the addition of VOT, this rose to 36% (18% improvement); with the addition of f0, this rose to 42% (6% improvement); and with the addition of vowel duration, this rose to 44% (2% improvement). Thus, the best model accounted for less than half of the variance in ratings, and included an acoustic property, f0, that was not predicted to play a role.

Similar models were created for stimuli with voiceless stop targets only; full statistical details are given in Table 3 in the Appendix. In the first model, VOT alone explained 40% of the variance in the ratings of perceived foreign accent—nearly as much as all four acoustic cues combined in the model for voiced targets. The addition of vowel quality brought this value to 45% (5% improvement), and the model that also included f0 accounted for 48% of the variance (3% improvement). Vowel duration, although significant in the stepwise model, did not contribute significantly to the explicitly specified regression model.

4.2 Overall rating patterns

Perceived foreign accent ratings by talker are shown for stimuli with voiced and voiceless stop targets separately in Figure 7. Separate one-way repeated measures ANOVAs with listeners as subjects and L1 as a within-subjects factor revealed significant effects of L1 on perceived foreign accent ratings for stimuli with voiced ($F(3,104) = 8.93$; $p < 0.001$) and voiceless ($F(3,104) = 10.29$; $p < 0.001$) stop targets. Post-hoc Bonferroni-corrected paired t-tests showed differences between all pairwise comparisons for voiced targets ($p < 0.00017$), as well as for voiceless targets ($p < 0.00017$). While there is quite a range of ratings within each L1 group, general patterns do emerge. As predicted, productions by L1 American English talkers were rated as having very little foreign accent, and productions by non-native talkers were rated as having greater degrees of foreign accent. In addition, productions by L1 Hindi talkers were rated as having the highest degree of foreign accent. While there was no basis for predicting a difference between perceived foreign accent ratings for productions by L1 Korean talkers as compared to productions by L1 Mandarin talkers, the latter received higher ratings, as shown in Figure 7.

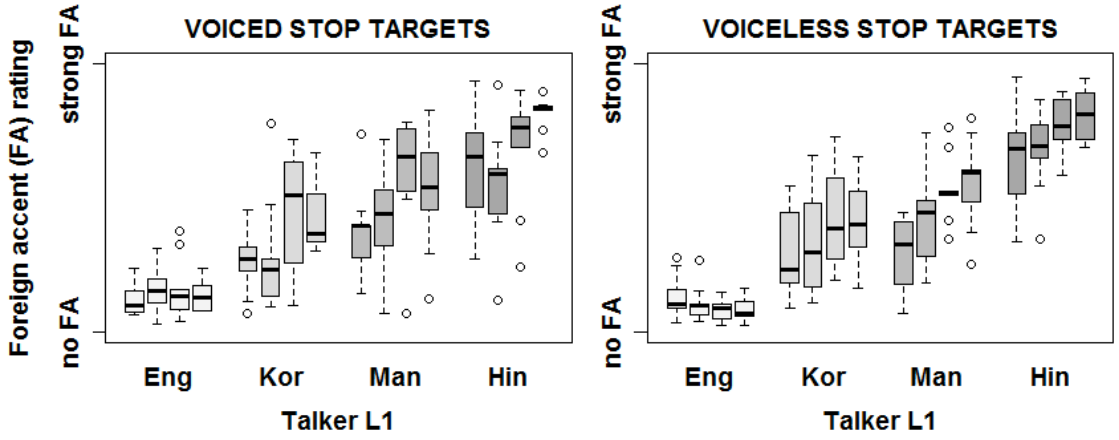

Figure 7. Ratings of perceived foreign accent by target voicing, talker L1, and talker identity

5 Discussion

The first goal of this research was to identify potential acoustic cues to perceived foreign accent. Acoustic properties were measured in very short stop-vowel sequences extracted from sentences in English. The same sequences were used as stimuli in a rating task, and the perceived foreign accent ratings were shown to be correlated with VOT, vowel quality, f0, and vowel duration for at least some of the stimuli. For stimuli with voiced stop targets, the best model included all four acoustic properties and accounted for 44% of the variance in perceived foreign accent ratings. For stimuli with voiceless stop targets, VOT, vowel quality, and f0 accounted for 48% of the variance in perceived foreign accent ratings, with VOT playing a substantial role.

The present study confirmed previous findings that VOT and vowel quality seem to be involved in the perception of foreign accent, suggested that f0 and vowel duration may also play minor roles, and showed that the relative importance of these cues differs somewhat for stimuli containing voiced stops targets as opposed to voiceless stop targets. This last fact is particularly notable given the frequent use of sentence-length stimuli, which contain many segments that may or may not be controlled for voicing, in studies of foreign accent. It remains to be seen whether such differences in relative importance persist in the perception of longer stimuli, or whether segmental cues are washed out by more global prosodic properties.

Although VOT was a significant predictor regardless of the subset of data analyzed, it clearly played a larger role in ratings of stimuli with voiceless stop targets than in ratings of stimuli with voiced stop targets. Figure 2 showed that a number of the voiced stop productions, especially by L1 Hindi talkers, were prevoiced, while voiceless stop productions were characterized by lag voicing. The difference in the contribution of VOT might have resulted from the listener population in this study; American English-speaking listeners are likely to be quite sensitive to lag voicing differences, as most word-initial productions by native speakers of American English are in this range. However, this effect might equally have been a general consequence of perception, as even listeners of languages with prevoiced stops seem to be more sensitive to the presence or absence of prevoicing than to the amount of prevoicing (van Alphen & McQueen, 2006). Thus, a convincing linear relationship between VOT and perceived foreign accent ratings might be unlikely for any stimuli with voiced stop targets. If so, a linear regression analysis would not be expected to perform well on this portion of the data. As previous investigations have generally focused only on voiceless stop targets, this complication has not been widely recognized.

While all four acoustic cues measured were found to contribute significantly to the model for at least one of the two target voicing categories, collectively they accounted for around half of the variance in perceived foreign accent ratings, leaving half of the variance unexplained. Stop-vowel stimuli were chosen so as to limit the number of acoustic cues listeners were able to attend to in performing the task, but it seems that listeners used more than the four acoustic cues measured here. As an individual's L1 is expected to influence L2 production, additional acoustic properties to consider in future analyses might be those that distinguish stops in the non-native talkers' L1s. For instance, voice quality and noise offset time are often invoked in phonetic descriptions of Hindi stops (Davis, 1994), and in Korean, voice quality and f0 at the onset of the following vowel are relevant to the 3-way stop distinction (Kong, Beckman, & Edwards,

2011). Although such cues are not needed for phonemic distinctions in American English, this does not mean that they were not perceived by American English-speaking listeners. If detectable in the stimuli, such cues might clearly identify certain productions as foreign and augment their ratings of perceived foreign accent.

The second goal of this research was to investigate whether some varieties of non-native English sound generally more accented than others. The large ranges of ratings for productions by talkers of each L1 indicate that listeners were not simply categorizing the productions into four discrete piles on the scale. Acoustic measurements differ among productions by talkers of different L1s, but also among productions by talkers from the same L1 and even among productions by the same talker. These multiple levels of variation might explain why f0 turned out to contribute to the regression models for both voiced and voiceless stop targets despite there being no effect of talker L1 on its difference values. Nonetheless, it seems clear that not all L1 backgrounds are equal when it comes to the perception of foreign accent. While productions by all non-native talkers were judged as sounding foreign-accented to some degree, those by L1 Hindi talkers were most accented, while those by L1 Mandarin talkers were more accented than those by L1 Korean talkers. These results suggest that listeners have some quantifiable concept of "foreign accent" more abstract than the "Brazilian Portuguese accent" or "Japanese accent" concepts that might have been explored in earlier studies, though the relationship between the abstract "foreign accent" and the various more specific incarnations is yet unclear.

These two sets of results also can be related to each other. For instance, the acoustic analysis showed that productions by L1 Hindi talkers differed significantly from productions by native talkers in their VOT values. Thus, there is a clear relationship between the importance of VOT as an acoustic cue and the high perceived foreign accent ratings for productions by L1 Hindi talkers. The directionality, however, is uncertain. Are productions by L1 Hindi talkers rated as most foreign-accented because they differ in VOT, or does VOT matter because English produced by L1 Hindi talkers is thought to sound highly foreign-accented? When asked if they could identify the accents present in the stimuli, 19 of 28 listeners identified "Indian" (15), "Hindi" (3), or "South Asian" (1), indicating that this variety of speech may have been easily identifiable. However, 18 of 28 listeners identified "Chinese" (9), "Mandarin" (1), "Asian" (7), or "East Asian" (1), while this variety of speech was rated as sounding significantly less foreign-accented. Only 2 of 28 listeners specifically identified "Korean," although some of the "Asian" and "East Asian" responses may have been at least partly in response to the productions by L1 Korean talkers. Further research is needed to investigate the role that stereotypes and familiarity with particular non-native varieties might play in the perception of foreign accent.

The results of this study are particularly interesting in light of the sociolinguistic situation of English in India, where "it is the 'associate official' language of the country and it also serves as a link language between the educated" (Gargesh, 2004, p. 992). Although they did not claim English as a native language, the L1 Hindi talkers in this study reported that they first began studying English between ages 3 and 4. In contrast, the L1 Korean and L1 Mandarin talkers in this study reported that they did not begin to learn English until at least age 11. Compared to the other non-native groups, the L1 Hindi talkers' experience with English started earlier and lasted longer. Their target variety of English was Indian rather than American, however, and the ratings from this

experiment suggest that this variety can sound quite foreign-accented to American English-speaking listeners. Thus, judgments about foreign accent are critically shaped by the listener's ideas about the world.

A listener's impressions regarding foreign accent develop over the course of day-to-day language use. However, most communication requires units much more complex than the stop-vowel sequences used as stimuli in this study, and the cues discovered here would not necessarily play the same roles in ratings of larger units like words or sentences. As longer stimuli would exhibit a much wider variety of potential acoustic cues, a considerably more complicated relationship between perceived foreign accent ratings and acoustic cues would seem inevitable. Nonetheless, the present work has taken the initial steps in addressing the question of what aspect of the speech signal causes listeners to perceive a foreign accent.

References

Bamford, John, and Ian Wilson. 1979. Methodological considerations and practical aspects of the BKB sentence lists. In *Speech-hearing tests and the spoken language of hearing-impaired children*, ed. by John Bench and John Bamford, 148-187. London: Academic Press.

Boersma, Paul, and David Weenink. 2012. *Praat: Doing phonetics by computer* (Version 5.3) [Computer program]. <http://www.praat.org/>.

Davis, Katharine. 1994. Stop voicing in Hindi. *Journal of Phonetics* 22.177-193.

Flege, James Emil. 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76.692-707.

Flege, James Emil, Murray Munro, and Ian MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America* 97.3125-3134.

Gargesh, Ravinder. 2004. Indian English: Phonology. In *A Handbook of varieties of English*, ed. by Edgar W. Schneider, Kate Burridge, Bernd Kortmann, Rajend Mesthrie, and Clive Upton, 992-1002. Berlin: Mouton de Gruyter.

Gluszek, Agata, and John F. Dovidio. 2010. The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review* 14(2).214-237.

Gonzalez-Bueno, Manuela. 1997. Voice onset time in the perception of foreign accent by native listeners of Spanish. *International Review of Applied Linguistics in Language Teaching* 35(4).251-262.

Hardman, Jocelyn B. 2010. The intelligibility of Chinese-accented English to international and American students at a US university. Columbus, OH: OSU dissertation.

Kong, Eun Jong, Mary E. Beckman, and Jan Edwards. 2011. Why are Korean tense stops acquired so early: The role of acoustic properties. *Journal of Phonetics* 39(2).196-211.

Ladefoged, Peter. 1999. American English. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, 41-44. Cambridge: Cambridge University Press.

Lado, Robert. 1957. *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press: Ann Arbor.

Lee, Hyun Bok. 1999. Korean. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, 120-123. Cambridge: Cambridge University Press.

Lee, Wai-Sum, and Zee, Eric. 2003. Illustrations of the IPA: Standard Chinese (Beijing). *Journal of the International Phonetic Association* 33(1).109-112.

Major, Roy C. 1987. English voiceless stop production by speakers of Brazilian Portuguese. *Journal of Phonetics* 15.197-202.

Munro, Murray J., & Tracey M. Derwing. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech. *Studies in Second Language Acquisition* 23.451-468.

Ohala, Manjari. 1999. Hindi. In *Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet*, 100-103. Cambridge: Cambridge University Press.

Oyama, Susan. 1976. A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5.261-285.

Riney, Timothy J., and Naoyuki Takagi. 1999. Global foreign accent and voice onset time among Japanese EFL speakers. *Language Learning* 49(2).275-302.

Riney, Timothy J., Naoyuki Takagi, and Kumiko Inutsuka. 2005. Phonetic parameters and perceptual judgments of accent in English by American and Japanese listeners. *TESOL Quarterly* 39(3).441-466.

Rubin, Donald L. 1992. Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants. *Research in Higher Education* 33(4).511-531.

van Alphen, Petra M., and James M. McQueen. 2006. The effect of Voice Onset Time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance* 32(1).178-196.

Appendix

Table 1. Stop-vowel sequences used as audio stimuli

Sequence	*Word*
/bæ/	back
/bʌ/	buckets
/bɑ/	boxes
/dɪ/	dish
/dɪ/	dinner
/dɔ/	dog
/gɚ/	girl
/gʊ/	good
/gɑ/	got
/pi/	people
/pɪ/	picture
/pæ/	packed
/teɪ/	table
/tu/	two
/tɑʊ/	towel
/kɪ/	kitchen
/keɪ/	came
/koʊ/	coat

Table 2. Regression models (stimuli with voiced stop targets)

Variable	*Value*	*Statistical test*
Model 1	$R^2 = 0.18$	$F(1,142) = 32.67$, $p < 0.001$
Vowel quality	$m_1 = 0.09$	$t(142) = 5.72$, $p < 0.001$
Model 2	$R^2 = 0.36$	$F(2,141) = 41.06$, $p < 0.001$
Vowel quality	$m_1 = 0.10$	$t(141) = 6.72$, $p < 0.001$
VOT	$m_2 = 0.32$	$t(141) = 6.36$, $p < 0.001$
Model 3	$R^2 = 0.42$	$F(3,140) = 35.93$, $p < 0.001$
Vowel quality	$m_1 = 0.09$	$t(140) = 6.04$, $p < 0.001$
VOT	$m_2 = 0.28$	$t(140) = 5.78$, $p < 0.001$
f0	$m_3 = 0.28$	$t(140) = 4.07$, $p < 0.001$
Model 4	$R^2 = 0.44$	$F(4,139) = 28.91$, $p < 0.001$
Vowel quality	$m_1 = 0.08$	$t(139) = 5.64$, $p < 0.001$
VOT	$m_2 = 0.28$	$t(139) = 5.90$, $p < 0.001$
f0	$m_3 = 0.27$	$t(139) = 3.99$, $p < 0.001$
Vowel duration	$m_4 = 0.20$	$t(139) = 2.21$, $p < 0.05$

Table 3. Regression models (stimuli with voiceless stop targets)

Variable	*Value*	*Statistical test*
Model 1	$R^2 = 0.40$	$F(1,142) = 96.21$, $p < 0.001$
VOT	$m_1 = 0.85$	$t(142) = 9.81$, $p < 0.001$
Model 2	$R^2 = 0.45$	$F(2,141) = 60.02$, $p < 0.001$
VOT	$m_1 = 0.76$	$t(141) = 8.82$, $p < 0.001$
Vowel quality	$m_2 = 0.04$	$t(141) = 3.82$, $p < 0.001$
Model 3	$R^2 = 0.48$	$F(3,140) = 44.91$, $p < 0.001$
VOT	$m_1 = 0.73$	$t(140) = 8.69$, $p < 0.001$
Vowel quality	$m_2 = 0.03$	$t(140) = 3.35$, $p < 0.01$
f0	$m_3 = 0.17$	$t(140) = 2.90$, $p < 0.01$
Model 4	$R^2 = 0.48$	$F(4,139) = 34.57$, $p < 0.001$
VOT	$m_1 = 0.71$	$t(139) = 8.25$, $p < 0.001$
Vowel quality	$m_2 = 0.03$	$t(139) = 2.66$, $p < 0.01$
f0	$m_3 = 0.17$	$t(139) = 2.98$, $p < 0.01$
Vowel duration	$m_4 = 0.11$	$t(139) = 1.51$, n.s.

AN INTRODUCTION TO RANDOM PROCESSES FOR THE SPECTRAL ANALYSIS OF SPEECH DATA

Patrick F. Reidy
Ohio State University

Abstract

Spectral analysis of acoustic data is a common analytical technique with which phoneticians have ample practical experience. The primary goal of this paper is to introduce to the phonetician, whose primary interest is the analysis of linguistic data, a portion of the theory of random processes and the estimation of their spectra, knowledge of which bears directly on the choices made in the process of analyzing time series data, such as an acoustic waveform. The paper begins by motivating the use of random processes as a model for acoustic speech data, and then introduce the spectral representation (or, spectrum) of a random process, taking care to relate this notion of spectrum to one that is more familiar to phoneticians and speech scientists. A final section presents two methods for estimating the values of the spectrum of a random process. Specifically, it compares the commonly-used (windowed) periodogram to the multitaper spectrum, and it is shown that the latter has many beneficial theoretical properties over the former.

Patrick F. Reidy. An introduction to random processes for the spectral analysis of speech data. *OSUWPL* Volume 60, Spring 2013, pp. 67-116.

1 Introduction

This paper discusses some of the statistical methods involved in the spectral analysis of speech data. Specifically, its aim is to introduce phoneticians to random processes, the class of mathematical object used to model speech data; their spectral representation; and some of the methods for estimating the values of a random process's spectrum.

In order to appreciate the place that random processes hold in a spectral analysis of speech data, we first consider a concrete example of such an analysis. Suppose that a researcher wishes to investigate the spectral properties of the English voiceless sibilant /s/. The first step in this investigation is to collect data by recording several tokens of /s/ from multiple English speakers. We refer to this type of data as *speech data*, measurements of the air pressure fluctuations caused by a particular speech sound wave, as sensed by a microphone. In practice, these measurements are typically stored by a digital recording device as a sequence of numbers, where each number represents the instantaneous air pressure at a given time.

So, the actual physical sound wave generated during speech production and the experimenter's record of that sound wave differ in basic ways. Whereas, the physical sound wave causes continuous air pressure fluctuations over a continuous interval of time, the record of the sound wave has been both sampled and quantized, which results in discrete air pressure fluctuations that occur over a discrete time interval. Because the researcher has access to only the record of the sound wave and because our focus is the analysis of speech data, we choose to represent a sound wave and its waveform as a numeric sequence.

Figure 1 shows the waveform of a token of /s/ that might be recorded by the researcher. The values of this waveform appear to vary randomly from one sample to the next. This random variation is expected in the waveform of /s/ because its noise source is generated by turbulent airflow, which by definition involves random air pressure variation. However, there are more fundamental sources of randomness in all speech data that affect not only turbulent sounds such as sibilants, but also quasi-periodic sounds like vowels.

One source of this randomness is the recording equipment itself. The microphone, by its very nature as a physical sensor, is subject to small random changes in its behavior over time. Since the microphone mediates the physical sound wave and its record, these small random changes in the microphone's behavior engender random errors in the recorded data. Likewise, the recording device may introduce low-frequency background noise whose intensity varies randomly over time.

Moreover, it is known that speech is subject to intra-speaker variation. The waveforms of two tokens of the same word spoken by the same person, even in proximate succession, are assured to show unpredictable differences in their values, which may be due to differences in speaking rate, vocal effort, articulatory gestures, etc. that the speaker, much less the researcher, is unable to control from one production to the next; hence, this variation can be considered random.

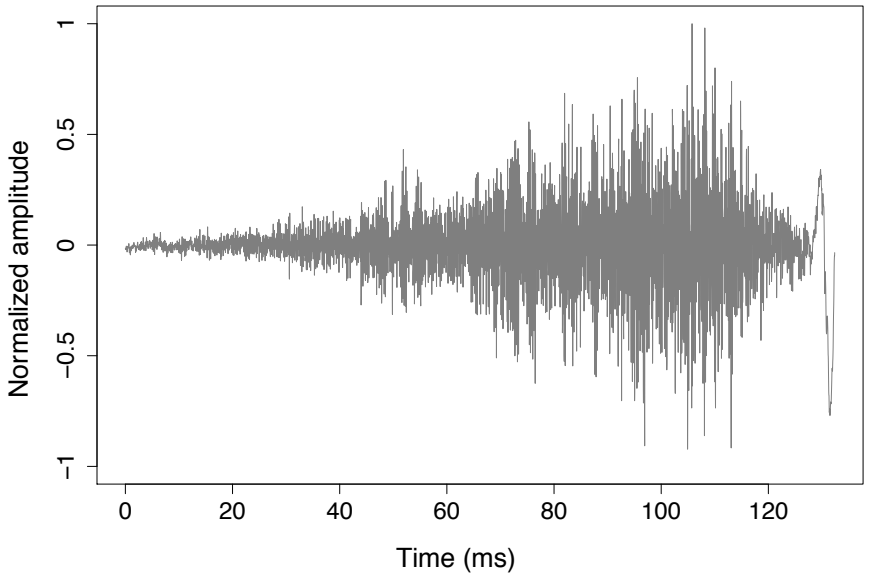

Figure 1: A realization of word-initial English /s/ excised from a token of 'sodas' as produced by an adult male speaker.

Due to the randomness intrinsic to speech data, each value of a waveform should be construed as a particular value taken by a random variable. For example, suppose that the researcher records a token of /s/ as the sequence of n numbers $x_1, x_2, \ldots, x_n$, where each value x_t is the t^{th} sampled value of the sound wave. Then, a natural model for the waveform of /s/ is a sequence of random variables $X_1, X_2, X_3, \ldots$, which just so happened to assume the values $x_1, x_2, \ldots, x_n$ when that particular token of /s/ was recorded. The decision to model the waveform of /s/ as a sequence of random variables correctly captures the fact that the values of the waveform of a token of /s/ are random in the sense discussed in the preceding two paragraphs, and motivates the introduction and definition of random processes.

Definition 1.1 (Random process)**.** A *random process* is a sequence of random variables, denoted by $\{X_t\}$, that are all defined on the same probability space, take values in the same measurable space, and are indexed by a variable t that ranges over (a subset of) the integers.

The measurable space in which each random variable takes its values is called the *state space* of the random process.

The linguistic objects suitable to be modeled by random processes are not limited to just the waveforms of phonetic segments. Indeed, the definition of a random process is a sequence of random variables, all of which are defined on a common probability space and share a common state space. The specifics of the probability space and state space are left open. So, all of the following could equally well be modeled by a random process: the waveform of a word, an f_0 track, a sequence of articulator positions, a text corpus. Each of

these examples reflects a change in the state space.

When a finite number of the variables in a random process $\{X_t\}$ assume values, the result is a sequence of numbers, referred to as a *realization* of the process and denoted by $\{x_t\}$; hence, the acoustic tokens of /s/ that form the data in the example are modeled as realizations of the random process that models /s/. So, when a random process's realizations are acoustic data, it should be clear that the random process models some waveform, not some sequence of states that describe constrictions in the vocal tract during the generation of that waveform, or some sequence of articulator postures that formed those vocal tract constrictions, or some sequence of motor unit activations that postured the articulators, or any other sequence of "articulatory states" at an even earlier point in the speech chain.

When a random process $\{X_t\}$ is used to model an acoustic waveform that has been sampled, the index t represents the (discrete) ordinal points in time at which the sound wave is sampled; hence, in the example of /s/, each random variable X_t models the t^{th} value of /s/'s waveform when sampled. If the sampling period T is known in seconds, then the "time" of each random variable in the random process can be given a physical meaning by associating each random variable X_t to the time tT seconds.

Once the researcher has collected a number of /s/ tokens, the spectral analysis can begin. For concreteness, suppose that the goal of the spectral analysis is to determine the peak frequency of the spectrum of /s/. From a procedural point of view, this analysis is straightforward: First, the spectrum of each /s/ token is computed; then, the peak frequency of each spectrum is determined; and finally, these values are used to estimate the peak frequency of /s/'s spectrum.

From a conceptual point of view, however, some elaboration is needed before this type of analysis can be considered meaningful. First, the notions of the "spectrum of /s/" and "the spectrum of a token of /s/" need to be clarified. Each of these ideas is resolved through the mathematical model of each linguistic object: A token of /s/ is modeled as a sequence of numbers, whose spectrum is known from the discrete Fourier transform (DFT). Therefore, the spectrum of a token of /s/ can be understood as the spectrum of the numeric sequence that models that /s/ token.

Likewise, the spectrum of /s/ refers to the spectrum of the random process that models /s/. But since the reach of traditional Fourier theory does not extend to random processes, this immediately exposes a hole in the logic of the procedure above. That is, the DFT is a map whose domain is a particular class of numeric sequence. This domain excludes all random processes; therefore, the DFT cannot be used to transform a random process into its spectrum. The change in mathematical object, whose spectrum is to be found, demands an extension of traditional Fourier theory. Without a theory of the spectral representation of random processes, a spectral analysis of /s/, or any other phonetic segment for that matter, is devoid of meaning.

In §2, the necessary extensions to traditional Fourier theory are reviewed. Specifically, it turns out that not all random processes have a spectral representation, so conditions on a

random process that guarantee the existence of its spectrum are presented. Furthermore, it is shown that each value of the spectrum of a random process $\{X_t\}$ depends on the infinite number of random variables in the process. However, $\{X_t\}$ is only ever observed as a realization of a finite number of variables; hence, each value of $\{X_t\}$'s spectrum can never be computed exactly. Instead, these values must be estimated from a finite realization.

In §3, two methods are presented for estimating the spectrum of a random process $\{X_t\}$ from a particular realization $x_1, x_2, \ldots, x_n$. Each method of estimation is evaluated analytically in order to explore how "close" to the true spectrum of $\{X_t\}$ an estimate computed from either method is expected to be. The form of these spectral estimators reveals the connection between the spectrum of $x_1, x_2, \ldots, x_n$ provided by the DFT and the spectrum of $\{X_t\}$. Specifically, both methods for estimating the spectrum of $\{X_t\}$ from $x_1, x_2, \ldots, x_n$ are based on the DFT of $x_1, x_2, \ldots, x_n$. This discussion, by extension, elucidates how the spectrum of a token of /s/ may be considered a representation of the spectrum of /s/.

Since the ultimate goal of the spectral analysis described above is the estimation of the peak frequency of /s/, rather than just its spectrum, the relationship between this or any other spectral property of /s/ and an estimate of it from a token of /s/ should be clarified as well. Mathematically, a spectral property of /s/ corresponds to a transformation of the spectrum of a random process $\{X_t\}$. For example, if the spectrum of $\{X_t\}$ is denoted by f_X, which ranges over a variable ω that denotes frequency, then the peak frequency of /s/ is given by the transformation

$$\mathsf{Peak}(X) = \arg\max_{\omega} f_X(\omega).$$

Similarly, a spectral property of a token of /s/ corresponds to a transformation of a spectral estimate computed from a realization $x_1, x_2, \ldots, x_n$. If this spectral estimate is denoted by S_x, which ranges over the discrete variable ω_j, then the peak frequency of the /s/ token is given by

$$\mathsf{Peak}(x) = \arg\max_{\omega_j} S_x(\omega_j).$$

While the discussion in §3 tells how S_x relates to f_X, it says nothing about how $\mathsf{Peak}(x)$ relates to $\mathsf{Peak}(X)$. The paper concludes with a discussion of the difficulties attendant with determining analytically how a spectral property of a random process $\{X_t\}$ relates to an estimate of that property computed from a realization $x_1, x_2, \ldots, x_n$. This difficulty of analysis implies that an analytic comparison of different methods for estimating a spectral property is for all practical purposes intractable. Instead, the researcher must justify which method of spectral estimation yields the "best" estimate of a given spectral property, by way of simulation rather than assuming that the relative merits of one spectral estimator over another transfer to estimates of spectral properties derived from that estimator.

2 Spectral representation of a random process

In this section, the theory of spectral representation for random processes is reviewed. The discussion is based on Shumway and Stoffer (2006), and the reader is referred there for a thorough general introduction to random processes and their spectral representation.

In the sequel, upper case letters $X, Y, \ldots$ are used to denote random variables; $\mathbb{E}(X)$ denotes the expected value of the random variable X; $\mathrm{Var}(X)$ denotes the variance of X; and $\mathrm{Cov}(X, Y)$ denotes the covariance between the random variables X and Y.[1] It is assumed that the reader is familiar with the meaning of all these terms.

In general, the methods from classical statistics cannot be applied to a random process because these methods assume that the random variables $\{X_t\}$ are independent and all follow the same distribution; however, a random process will not always obtain both of these properties. When used to model the waveform of a phonetic segment, the dependence structure of a random process and the change in its distributional properties over time are due to the nature of and physical constraints on speech production. First, speech production necessarily involves the movement of articulators, and as the posture of the articulators changes over time, the generated sound wave changes as well; hence, the distributional properties of a random process that models the wave form are expected to change with time as well. Second, the articulators move smoothly during the production of speech, and the posture that they can assume next depends on their current postural state. This dependence is projected forward in the speech chain, to the acoustic sound wave.

A complete description of the dependence structure of a random process $\{X_t\}$ would be had from knowing the joint cumulative distribution function of all finite subsets of random variables in $\{X_t\}$; however, such a complete description is usually unattainable. Instead, a much more limited description of $\{X_t\}$'s dependence structure is taken from its autocovariance function, which reports the covariance between each pair of variables in $\{X\}$. Below it is shown that the autocovariance function is intimately related to the spectral representation of a random process.

Definition 2.1 (Autocovariance function)**.** If $\{X_t\}$ is a random process, then the *autocovariance function* γ_X is defined by

$$\gamma_X(s, t) = \mathrm{Cov}(X_s, X_t). \tag{1}$$

In general, a process does not have a spectral representation; however, there is a very general subclass of random processes—the (weakly) stationary processes—which do admit such a frequency-domain representation.

[1] In general, it is possible that any of $\mathbb{E}(X)$, $\mathrm{Var}(X)$, or $\mathrm{Cov}(X, Y)$ may not converge to a value, making that value undefined; however, in the sequel it is assumed that all random variables have a finite expected value and variance and that all pairs of random variables have a finite covariance.

Definition 2.2 (Stationary process)**.** A random process $\{X_t\}$ is said to be *(weakly) stationary*[2] if it satisfies the following conditions:

1. $\mathbb{E}(X_t) = \mu$, for all t in the index set;
2. $\mathrm{Var}(X_t) < \infty$, for all t in the index set;
3. $\mathrm{Cov}(X_s, X_t) = \mathrm{Cov}(X_{s+h}, X_{t+h})$, for all s, t, $s+h$, and $t+h$ in the index set.

A process that does not satisfy the three conditions above is said to be *non-stationary*.

The third condition says that the covariance between any two random variables in a stationary process depends only on the amount of time that separates them, which allows its autocovariance function to be expressed in terms of a single variable denoting the separation between two random variables in the process: If $\{X_t\}$ is a stationary process with autocovariance function γ_X, then for all indices s and t with $h = s - t$, it follows that

$$\begin{aligned}\gamma_X(s,t) &= \gamma_X(s+h, t+h) \\ &= \gamma_X(s+h, s) \\ &= \gamma_X(h, 0),\end{aligned}$$

which does not depend on either time argument s or t. Hence, the autocovariance function of a stationary process can be expressed as a function of just the separation (or *lag*) h between two random variables,

$$\gamma_X(h) =_{def} \mathrm{Cov}(X_0, X_h). \tag{2}$$

If the $\{X_t\}$ models a waveform that is sampled with sampling period T seconds, then the lag h that separates two random variables in the process can be given the physical meaning of a separation of hT seconds.

The spectral representation of a stationary process can now be introduced. It can be proved that any stationary process can be expressed as a random linear combination of simple periodic functions oscillating at different frequencies (Shumway and Stoffer, 2006, Theorem C.2). Additionally, the autocovariance function of a stationary process also has a spectral representation, which is provided by the following theorem, stated without proof (Shumway and Stoffer, 2006, Property P4.1 & Theorem C.3).

Theorem 2.3 (Spectral Representation)**.** *If $\{X_t\}$ is a stationary process whose autocovariance function γ_X satisfies*

$$\sum_{h=-\infty}^{\infty} |\gamma_X(h)| < \infty,$$

[2]By contrast, a process is said to be *strictly stationary* if the distributional properties of all finite subcollections of random variables in the process do not depend on time.

then there is a unique function f_X for which

$$\gamma_X(h) = \int_{-1/2}^{1/2} f_X(\omega) e^{2\pi i \omega h} \, \mathrm{d}\omega, \qquad h = 0, \pm 1, \pm 2, \ldots \tag{3}$$

The function f_X in (3) is called the spectral density *or* spectrum *of $\{X_t\}$ and is defined by*

$$f_X(\omega) = \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{-2\pi i \omega h}, \qquad \omega \in \mathbb{R}. \tag{4}$$

Readers who are familiar with traditional Fourier theory may notice that the spectral density f_X above is the Fourier transform of the (aperiodic, discrete) autocovariance function γ_X (Beerends et al., 2003, §18.5). This implies that f_X and γ_X uniquely determine each other, and that the spectral density f_X and the autocovariance function γ_X contain the same information since each value of γ_X can be recovered from f_X by integrating the right-hand side of (3). Therefore, we take the spectral density f_X of a stationary process $\{X_t\}$ as its foremost spectral representation and in the remainder of this section present some of the practical consequences of Theorem 2.3 for the spectral analysis of speech data.

2.1 Existence of a spectral representation of speech data

The first of these consequences concerns the speech data that a researcher is able to use to investigate spectral properties of a phonetic segment's waveform. Recall the example from the introduction, in which the waveform of /s/ is modeled by a random process $\{X_t\}$, and the data used in the study are modeled as realizations of $\{X_t\}$. In this setting, Theorem 2.3 implies that it is only meaningful to talk about the spectrum of the waveform of /s/ if that waveform is stationary. If /s/'s waveform is not stationary, then a stationary portion of /s/ must be isolated and used for the purposes of the spectral analysis.

Since the waveform of /s/ is only ever observed through a realization of it, this condition on the existence of a spectrum of (a portion of) /s/'s waveform, this raises the question of how to determine whether a particular token of /s/ is a realization of a stationary process. A rough but common method for checking this involves plotting the recorded token $x_1, x_2, \ldots, x_n$ as a function of time, and visually inspecting the mean and variance properties of its waveform. In particular, if the data is a realization of a stationary process, then it follows from definition 2.2(1) that the mean of $x_1, x_2, \ldots, x_n$ should be constant across time, and it follows from the following proposition that the variance should be constant as well.

Proposition 2.4 (Variance of a stationary process)**.** *If $\{X_t\}$ is a stationary process, then for every X_s and X_t in the process,* $\mathrm{Var}(X_s) = \mathrm{Var}(X_t)$.

Proof. Let X_s and X_t be random variables from a stationary process, and let $h = t - s$. Then, $\mathrm{Var}(X_s) = \mathrm{Cov}(X_s, X_s) = \mathrm{Cov}(X_{s+h}, X_{s+h}) = \mathrm{Cov}(X_t, X_t) = \mathrm{Var}(X_t)$.

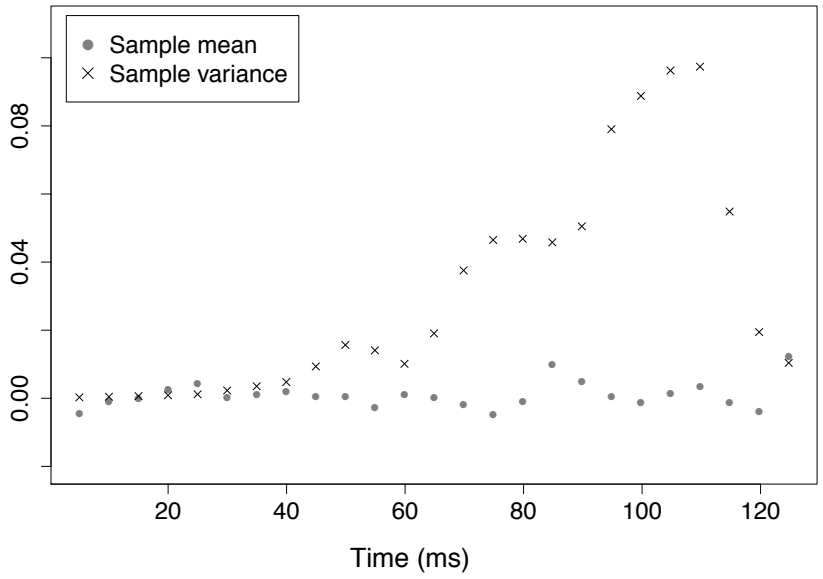

Figure 2: The sample mean (gray) and sample variance (black) of successive 10 ms data windows, with 5 ms overlap among adjacent windows, taken from the /s/ token shown in Figure 1. The value of each statistic is plotted against the time of the midpoint of the data window from which it was calculated.

The first and last equalities follow from the definition of covariance, and the second equality follows from the assumption that X_s and X_t come from a stationary process. □

Figure 2 shows the temporal progression of the sample mean and sample variance of successive 10 ms windows taken from the /s/ token shown in Figure 1. These statistics estimate the behavior of the evolution of the mean and variance of the random process $\{X_t\}$ that models /s/. From these plots, it is seen that the mean remains approximately constant, but the variance increases with time before decreasing sharply. So, this token of /s/ does not seem to be a realization of a stationary process, which, when considered in light of Theorem 2.3, suggests that it would be imprudent, much less meaningful, to use all the data from this token to estimate spectral properties of /s/.

In order to surmount this problem, the data can be used to hypothesize the location of a stationary subprocess of $\{X_t\}$, whose spectrum can be used as a proxy for that of the entire process. The data in Figure 2 suggest that the initial 40 ms interval of /s/ is stationary, as are the intervals between 40 and 60 ms and 70 and 90 ms. However, when automating a spectral analysis over a large data set, it is practically impossible to inspect each token individually in order to locate a stationary portion. Instead, it is common practice to take from each token a short interval placed in the same relative location, e.g. a 20 ms interval centered at the temporal midpoint of the waveform. It is taken on faith that the interval is of short enough duration that the statistical properties of the random process do not change

too drastically to violate the condition of stationarity. The fact that phoneticians typically restrict spectral analyses to "steady-state" portions of speech data suggests that random processes already occupy a very real, albeit unappreciated, role in phonetic analyses.

2.2 The domain of the spectral density function

In equation (4), the spectral density function f_X is defined over the entire real line; however, phoneticians are accustomed to visualizing the spectrum of speech data only on the interval of frequency values that ranges from 0 to the Nyquist frequency, $1/2T$ Hz, where T is the sampling period of the recorded data. Propositions 2.5 and 2.7 reconcile this discrepancy between theory and practice.

Proposition 2.5 (f_X is a periodic function)**.** *If $\{X_t\}$ is a random process that models an acoustic wave that is sampled with a sampling period T seconds, then its spectral density f_X is a periodic function with period $1/T$ Hz.*

Proof. From the discussion immediately following equation (2), each lag value h corresponds to hT units of time if $\{X_t\}$ models an acoustic wave that is sampled with sampling period T seconds. Therefore, equation (4) can be written as

$$f_X(\omega) = \sum_{h=-\infty}^{\infty} \gamma_X(hT) e^{-2\pi i \omega h T}, \tag{5}$$

where ω is expressed in Hz. Evaluating f_X at $\omega + 1/T$ then yields

$$\begin{aligned} f_X(\omega + 1/T) &= \sum_{h=-\infty}^{\infty} \gamma_X(hT) e^{-2\pi i(\omega + 1/T)hT} && (6)\\ &= \sum_{h=-\infty}^{\infty} \gamma_X(hT) e^{-2\pi i \omega h T} e^{-2\pi i h} && (7)\\ &= \sum_{h=-\infty}^{\infty} \gamma_X(hT) e^{-2\pi i \omega h T} && (8)\\ &= f_X(\omega). && (9) \end{aligned}$$

Equation (8) follows by virtue of the identity $e^{i\pi n} = 1$ for any integer n; in this case, $n = -2h$. □

The preceding proposition shows that the values of f_X are determined by the values that it takes on any interval whose size is $1/T$. Proposition 2.7 shows that the size of this interval can effectively be cut in half.

Lemma 2.6 (γ_X is an even function)**.** *If γ_X is the autocovariance function of a stationary process $\{X_t\}$ as defined by equation (2), then γ_X is an even function in the sense that $\gamma_X(-h) = \gamma_X(h)$ for all h.*

Proof. If γ_X is as described in the statement of the lemma, then

$$\begin{aligned}\gamma_X(h) &= \mathrm{Cov}(X_0, X_h) && (10)\\ &= \mathrm{Cov}(X_{-h}, X_0) && (11)\\ &= \mathrm{Cov}(X_0, X_{-h}) && (12)\\ &= \gamma_X(-h) && (13)\end{aligned}$$

Equation (11) follows from definition 2.2(3); equation (12), from the elementary fact that $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$ for all random variables X and Y. □

Proposition 2.7 (f_X is an even function)**.** *If f_X is the spectral density function of a stationary process as defined by equation (4), then f_X is an even function.*

Proof. If f_X is the spectral density function of a stationary process, then

$$f_X(-\omega) = \sum_{h=-\infty}^{\infty} \gamma_X(h) e^{2\pi i \omega h}. \tag{14}$$

Making the change of variable $h = -j$ yields

$$\begin{aligned}f_X(-\omega) &= \sum_{j=\infty}^{-\infty} \gamma_X(-j) e^{-2\pi i \omega j} && (15)\\ &= \sum_{j=\infty}^{-\infty} \gamma_X(j) e^{-2\pi i \omega j}. && (16)\end{aligned}$$

Equation (16) follows from Lemma 2.6. Comparison of equation (16) to equation (4) reveals that the former is just an alphabetic variant of the latter, where the summation is carried out in reverse. Therefore, it follows that $f_X(-\omega) = f_X(\omega)$, which proves the proposition. □

Taken together Propositions 2.5 and 2.7 show that if $\{X_t\}$ is a stationary process that models an acoustic wave sampled with sampling period T seconds, then its spectrum f_X is completely determined by the values that f_X takes on the frequency interval $[0, 1/2T]$ Hz. Since f_X is an even function, the values that it takes on the interval $[0, 1/2T]$ can be used to reconstruct its values on the interval $[-1/2T, 1/2T]$. The size of this reconstructed interval is $1/T$; hence, the values taken by f_X on this interval completely determine f_X on its entire domain since f_X is periodice with period $1/T$. Consequently, the spectrum of any given waveform need only be considered on the interval $[0, 1/2T]$, and in the following section all graphs of these spectra are shown only on this interval.

3 Spectral Estimation

In equation (4), each ordinate of the spectral density, $f_X(\omega)$, is expressed in terms of the autocovariance function γ_X; however, it is possible to express each ordinate in terms

of the random variables in $\{X_t\}$ by replacing γ_X in (4) with the righthand side of equation (2),

$$f_X(\omega) = \sum_{h=-\infty}^{\infty} \mathrm{Cov}(X_0, X_h) e^{-2\pi i \omega h}, \qquad \omega \in \mathbb{R}. \tag{17}$$

From this equation, it is immediately clear that the computation of each ordinate of f_X requires knowledge of the distributional properties of all the random variables in $\{X_t\}$; however, the random process is only ever observed as a finite realization $x_1, x_2, \ldots, x_n$. So, the value of each ordinate $f_X(\omega)$ cannot be computed exactly, but must instead be estimated.

A method for estimating the ordinates of f_X, which often takes the form of a function of a finite number of random variables $X_1, X_2, \ldots, X_n$ from a stationary process $\{X_t\}$, is referred to as a *spectral estimator*. The random variables $X_1, X_2, \ldots, X_n$ are called a *sample* of the process $\{X_t\}$. This section presents two spectral estimators that have been used in the spectral analysis of speech data: the windowed periodogram and the multitaper spectrum. Each of these spectral estimators finds its roots in the discrete Fourier transform (DFT), a spectral transform that is typically defined in terms of a finite numeric sequence (see Beerends et al. (2003, p. 360)). For the discussion of spectral estimators that follows, it is more convenient to define the DFT in terms of random variables rather fixed numbers.

Definition 3.1 (Discrete Fourier transform)**.** If $X_1, X_2, \ldots, X_n$ is a finite sequence of random variables from a stationary process $\{X_t\}$, then the *discrete Fourier transform* d_X of the sample is defined by

$$d_X(\omega_j) = \sum_{t=1}^{n} X_t e^{-2\pi i \omega_j t}, \tag{18}$$

where $\omega_j = j/n$ for $j = 0, \ldots, n-1$. The frequencies ω_j are referred to as the *Fourier frequencies*.

Other commonly encountered spectral transformations derived from the DFT are the *amplitude spectrum* $|d_X|$, defined by

$$|d_X|(\omega_j) = |d_X(\omega_j)|, \tag{19}$$

and the *power spectrum* $|d_X|^2$, defined by

$$|d_X|^2(\omega_j) = |d_X(\omega_j)|^2. \tag{20}$$

In equation (18), each ordinate of the DFT, $d_X(\omega_j)$ is defined as a sum of the random variables $X_1, X_2, \ldots, X_n$; hence, $d_X(\omega_j)$ is a univariate random variable since a sum of univariate random variables is itself a univariate random variable. Furthermore, each $d_X(\omega_j)$ estimates the value of $f_X(\omega_j)$, and as such is an example of a *point estimator*, i.e. an estimator of a single value. It follows that it is meaningful to investigate each ordinate's

distributional properties, such as its expected value and its variance. Knowlege of these properties for $d_X(\omega_j)$ enables a discussion of its *bias* and *mean square error (MSE)* as a point estimator. The latter is commonly used as a measure of the quality of a point estimator, so by extension the spectral estimators presented below can be compared via the MSE of their ordinates.

Definition 3.2 (Bias)**.** If $\hat{\theta}$ is a point estimator of a number θ, then the *bias* of $\hat{\theta}$, denoted $\beta\left(\hat{\theta}\right)$, is defined to be $\beta\left(\hat{\theta}\right) = \mathbb{E}\left(\hat{\theta}\right) - \theta$.

If $\beta\left(\hat{\theta}\right) = 0$, then $\hat{\theta}$ is said to be an *unbiased* estimator.

Definition 3.3 (Mean square error)**.** If $\hat{\theta}$ is a point estimator of a number θ, then the *mean square error* of $\hat{\theta}$, denoted $\mathsf{MSE}\left(\hat{\theta}\right)$, is defined to be $\mathsf{MSE}\left(\hat{\theta}\right) = \beta\left(\hat{\theta}\right) + \mathrm{Var}\left(\hat{\theta}\right)$.

The rest of this section is devoted to introducing and comparing the windowed periodogram and the multitaper spectrum. For each spectral estimator, its bias and variance are discussed only qualitatively; however, some comparison of the two estimators is still possible.

3.1 The windowed periodogram

The periodogram arises from scaling the power spectrum in (20) by the inverse of the number n of random variables available to the estimator.

Definition 3.4 (Periodogram)**.** If $X_1, X_2, \ldots, X_n$ are a sample from a stationary process $\{X_t\}$, then the *periodogram* I_X of the sample is defined by

$$I_X(\omega_j) = n^{-1}|d_X(\omega_j)|^2, \tag{21}$$

where j and ω_j are as they are in definition (3.1).

The periodogram is, in some sense, the most "direct" estimator of the spectral density f_X given a particular sample $X_1, X_2, \ldots, X_n$. To see why this is so, recall the definition of f_X from equation (4). One immediately apparent method for estimating $f_X(\omega)$ is to estimate the autocovariance function γ_X and then compute the DFT of the result. A common estimator of γ_X is the *sample autocovariance function* (Shumway and Stoffer, 2006, p. 30).

Definition 3.5 (Sample autocovariance function)**.** If $X_1, X_2, \ldots, X_n$ is a sample of a random process $\{X_t\}$, then the *sample autocovariance function* is defined by

$$\hat{\gamma}_X(h) = n^{-1}\sum_{t=1}^{n-h}(X_{t+h} - \bar{X})(X_t - \bar{X}), \tag{22}$$

where $\bar{X} = n^{-1}\sum_t X_t$ is called the *sample mean*.

It is possible to show that for Fourier frequencies other than $\omega_0 = 0$ the DFT of the sample autocovariance function is equal to the periodogram.

Proposition 3.6 (DFT of the sample autocovariance function)**.** *If $X_1, X_2, \ldots, X_n$ is a sample of a stationary process $\{X_t\}$, with sample autocovariance function $\hat{\gamma}_X$ and periodogram I_X, then for Fourier frequencies other than $\omega_0 = 0$,*

$$I_X(\omega_j) = \sum_{|h|<n} \hat{\gamma}_X(h) e^{-2\pi i \omega_j h}.$$

Proof. First note that for $\omega_j \neq 0$, the DFT can be written as[3]

$$d_X(\omega_j) = \sum_{t=1}^{n} (X_t - \bar{X}) e^{-2\pi i \omega_j t}, \tag{23}$$

Therefore, for Fourier frequencies other than ω_0 it follows that

$$I_X(\omega_j) = n^{-1}|d_X(\omega_j)|^2 = n^{-1} \sum_{t=1}^{n} \sum_{s=1}^{n} (X_t - \bar{X})(X_s - \bar{X}) e^{-2\pi i \omega_j (t-s)} \tag{24}$$

$$= n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X})(X_t - \bar{X}) e^{-2\pi i \omega_j h} \tag{25}$$

$$= \sum_{|h|<n} \hat{\gamma}_X(h) e^{-2\pi i \omega_j h}. \tag{26}$$

Comparing equation (26) to definition 3.1, it is clear that the periodogram is equal to the DFT of the sample autocovariance function. □

This proposition establishes a nice parallel among the spectral representations of a stationary process and a sample of that process: The spectrum of each is the Fourier transform of its appropriate autocovariance function. In order to establish a more direct relationship between the spectral density and an estimator of it, the *windowed periodogram* is introduced.

Definition 3.7 (Windowed periodogram)**.** If $X_1, X_2, \ldots, X_n$ is a sample of a stationary process, and $w_1, w_2, \ldots, w_n$ is a sequence of numbers, then the *w-windowed periodogram* I_{wX} of the sample is defined by

$$I_{wX}(\omega_j) = n^{-1} \sum_{t=1}^{n} w_t X_t e^{-2\pi i \omega_j t}. \tag{27}$$

The sequence of numbers $w_1, w_2, \ldots, w_n$ is referred to as a *data window* or *data taper*.

[3]For any complex number $z \neq 1$, $\sum_{t=1}^{n} z^t = z(1 - z^n)/(1 - z)$. Let $\omega_j \neq 0$ and let $z = \exp(-2\pi i \omega_j)$. Then, $z \neq 1$ and $z^n = \exp(-2\pi i \omega_j n) = \exp(-2\pi i j n/n) = \exp(-2\pi i j) = 1$ since j is an integer; hence, $\sum_{t=1}^{n} z^n$ = z(1-1)/(1-z) = 0.

Then, to prove equation (23), expand the sum therein and cancel the $\bar{X} \sum_{t=1}^{n} e^{-2\pi i \omega_j t}$ term.

In Proposition 3.8 below, it is assumed that $X_1, X_2, \ldots, X_n$ is a sample from a zero-mean process $\{X_t\}$, meaning that $\mathbb{E}(X_t) = 0$, for each random variable X_t in the process. It is likely that a zero-mean process is a valid model for the acoustic waveform of speech because the sound waves generated during speech production travel as a chain of increases and decreases in air pressure, which are likely to cancel each other over time. Indeed, the data shown in Figure 2 suggest that the acoustic waveform of /s/ is well-modeled by a zero-mean process.

Proposition 3.8 (Expected value of periodogram ordinates)**.** *If $X_1, X_2, \ldots, X_n$ is a sample of a zero-mean stationary process $\{X_t\}$, and $w_1, w_2, \ldots, w_n$ is a data window, then the expected value of the w-windowed periodogram I_{wX} is*

$$\mathbb{E}[I_{wX}(\omega_j)] = \int_{-1/2}^{1/2} W_n(\omega_j - \omega) f_X(\omega)\ \mathrm{d}\omega, \tag{28}$$

where

$$W_n(\omega) = n^{-1} \left| \sum_{t=1}^{n} w_t e^{-2\pi i \omega t} \right|^2, \quad w \in \mathbb{R}. \tag{29}$$

W_n is called the kernel *of the data window $w_1, w_2, \ldots, w_n$.*

Proof. If the righthand side of (27) is expanded and one of the variables of summation changed to $h = t - s$, the result is

$$I_{wX}(\omega_j) = n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} X_t X_{t+|h|} e^{-2\pi i \omega_j h}.$$

Taking the expectation of both sides yields

$$\mathbb{E}[I_{wX}(\omega_j)] = n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} e^{-2\pi i \omega_j h} \mathbb{E}(X_t X_{t+|h|}) \tag{30}$$

$$= n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} e^{-2\pi i \omega_j h} \mathbb{E}[(X_t - \mathbb{E}(X_t))(X_{t+|h|} - \mathbb{E}(X_{t+|h|}))] \tag{31}$$

$$= n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} e^{-2\pi i \omega_j h} \gamma_X(h), \tag{32}$$

where the first equation follows from the linearity of the expected value operator; the second, from the fact that $\{X_t\}$ is a zero-mean process; and the third from equation (2).

Finally, substituting the righthand side of (3) for $\gamma_X(h)$ gives

$$\mathbb{E}[I_{wX}(\omega_j)] = \int_{-1/2}^{1/2} n^{-1} \sum_{|h|<n} \sum_{t=1}^{n-|h|} w_t w_{t+|h|} e^{-2\pi i(\omega_j-\omega)h} f_X(\omega)\ \mathrm{d}\omega \tag{33}$$

$$= \int_{-1/2}^{1/2} n^{-1} \left| \sum_{t=1}^{n} w_t e^{-2\pi i(\omega_j-\omega)t} \right|^2 f_X(\omega)\ \mathrm{d}\omega \tag{34}$$

$$= \int_{-1/2}^{1/2} W_n(\omega_j - \omega) f_X(\omega)\ \mathrm{d}\omega, \tag{35}$$

where the last equation follows from (29). □

Proposition 3.8 shows that the w-windowed periodogram I_{wX} and the spectral density f_X are mediated by the kernel W_n of the particular data window used on the sample. More specifically, the integral on the righthand side of equation (35) says that the expected value of the estimator $I_{wX}(\omega_j)$ is found by taking the kernel W_n, "laying it on top" of f_X so that $W_n(0)$ coincides with $f_X(\omega_j)$, multiplying the values of each function that overlap, and then summing these products. This operation is called the *convolution* of W_n and f_X.

However, it is important to recognize that equation (35) does not describe how the windowed periodogram estimate of $f_X(\omega_j)$ is computed from a realization; that information is found in equation (27). Instead, equation (35) provides information about how the estimator $I_{wX}(\omega_j)$ would behave if a number of estimates of $f_X(\omega_j)$ were computed from different realizations and then averaged, which is a doorway to the bias of $I_{wX}(\omega_j)$.

3.1.1 Bias properties of the windowed periodogram

It should be clear from equation (35) that in order for the expected value of $I_{wX}(\omega_j)$ to be determined, it is necessary to know both the kernel W_n and the spectral density f_X; however, in applications involving speech data, it is rarely the case that anything is known about f_X since this would require knowledge of the distributional properties of the random process $\{X_t\}$ that models the speech data. It is possible that such distributional knowledge could become available from a complete theory of the aeroacoustics of speech production, but at the moment this theory is lacking. Consequently, the bias of each ordinate in the windowed periodogram, $\beta\left(I_{wX}(\omega_j)\right) = \mathbb{E}\left[I_{wX}(\omega_j)\right] - f_X(\omega_j)$, is unknown because both terms involved in its computation depend on f_X.

Since the direct computation of $\beta\left(I_{wX}(\omega_j)\right)$ is often impossible in practice, the bias properties of the windowed periodogram are explored through the kernel W_n, whose form depends on the particular window applied to the data before the spectral estimate is computed. This section discusses the kernel's of two data windows that should be familiar to phoneticians and other speech researchers: the rectangular window and the Hamming window.

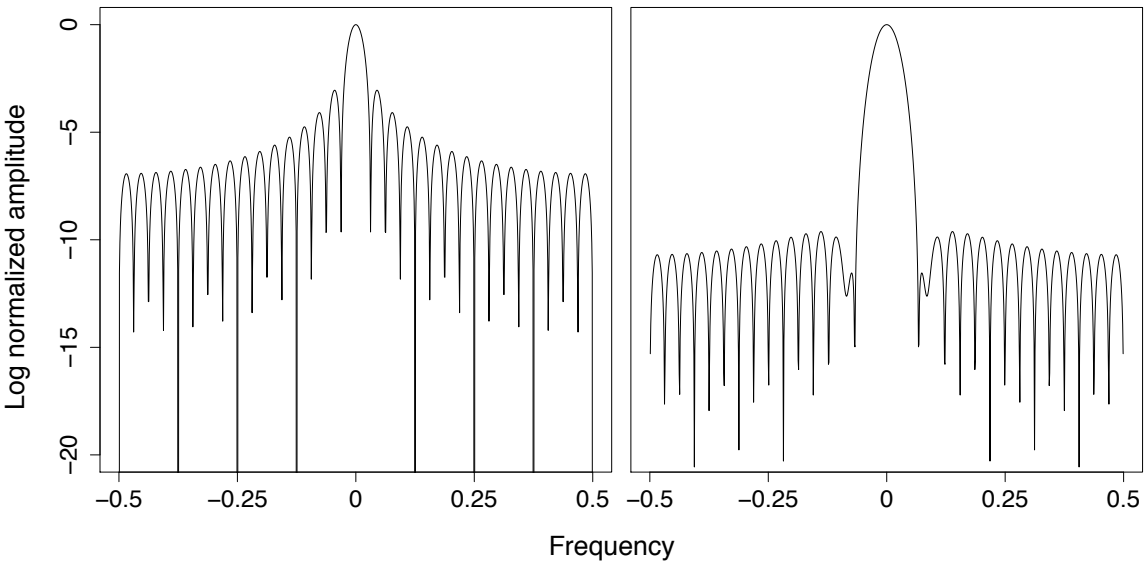

Figure 3: The kernel of the 32-point rectangular window (left panel) and the 32-point Hamming window (right panel). The values of each kernel were normalized by dividing by the maximum value of the kernel.

Definition 3.9 (Rectangular window)**.** The *(n-point) rectangular window* $r_1, r_2, \ldots, r_n$ is the sequence of n elements, each of which is equal to 1

Since $r_t X_t = X_t$ for $t = 1, 2, \ldots, n$, the rectangular window can be thought of as the "default" data window applied to the sample $X_1, X_2, \ldots, X_n$ when no other data window is used. From this it follows that the periodogram from Definition 3.4 is equal to the windowed periodogram from Definition 3.7 whose data window is the rectangular window; hence, the relationship between the windowed periodogram ordinate $I_{wX}(\omega_j)$ and the spectral density f_X established in Proposition 3.8 applies to the "unwindowed" periodogram as well, which implies that the bias properties of the periodram ordinates depend on the kernel of the rectangular window.

The kernel of the 32-point rectangular window is shown in the left panel of Figure 3. The shape of this kernel is characterised by a dominant peak, called the *main lobe*, centered at 0 with several other peaks, referred to collectively as the *side lobes*, on either side of it, whose respective heights decrease with their distance from the main lobe. While the number of side lobes in the kernel of a rectangular window depends on the length n of the window, the downward sloping pattern from the peak of the main lobe through the peaks of the side lobes is the same independent of n.

Consider how, according to equation (35), the shape of a kernel W_n affects the expected value, and by extension the bias, of $I_{wX}(\omega_j)$. The bias of $I_{wX}(\omega_j)$ is minimized when the

righthand side of this equation equals $f_X(\omega_j)$; however, when W_n and f_X are convolved, the value of f_X at each frequency $\omega \neq \omega_j$ is scaled by $W_n(\omega_j - \omega)$, and if $W_n(\omega_j - \omega) \neq 0$, then $W_n(\omega_j - \omega) f_X(\omega) \neq 0$ as well. Consequently, the degree to which $\mathbb{E}[I_{wX}(\omega_j)]$ is influenced by the value of the spectral density at a frequency $\omega \neq \omega_j$ is directly related to the magnitude of $W_n(\omega_j - \omega)$. Therefore, the height of the sidelobes of W_n gives some indication of the extent to which $\mathbb{E}[I_{wX}(\omega_j)]$ is corrupted by the values of f_X at frequencies different, and potentially far away, from $f_X(\omega_j)$, which in turn increases the magnitude of its bias. In sum, the height of the sidelobes of a kernel is a rough proxy measure of the bias of the spectral estimator related to that kernel—the greater the height of the kernel's sidelobes, the more biased the estimator.

The righthand panel of Figure 3 shows the kernel of the 32-point Hamming window.

Definition 3.10 (Hamming window)**.** The *n-point Hamming window* $h_1, h_2, \ldots, h_n$ is the sequence of numbers defined by

$$h_t = 0.5 \left(1 - \cos\left(\frac{2\pi(t-1)}{n-1}\right)\right), \quad t = 1, 2, \ldots, n. \tag{36}$$

The size of the sidelobes in the kernel of the Hamming window, relative to those in the rectangular window's kernel, suggests that the Hamming-windowed periodogram I_{hX} has better bias properties than the rectangular-window periodogram I_{rX}. Further support for this conclusion is provided by Figure 4, which shows a Hamming-window periodogram spectral estimate overlaid on a rectangular-window periodogram estimate, both of which were computed from the center 20 ms of the token of /s/ shown in Figure 1. Both estimates suggest that the most prominent peak of the spectral density occurs just below 5 kHz. Taking this together with Proposition 3.8 and both panels of Figure 3, it is expected that at high frequencies the values of the rectangular-windowed periodogram estimate would be higher than those of the hamming-windowed periodogram estimate since the sidelobes of the rectangular window's kernel are larger than those of the Hamming window's kernel.

While the differences in bias properties between the rectangular- and the Hamming-windowed periodogram are borne out by the example estimates in Figure 4, for the purposes of analyzing speech data, it is more important to focus on how or whether these differences affect the analysis rather than to focus on the purely theoretical concern as to whether they exist at all. For example, both windowed periodogram estimates share roughly the same shape; hence, if the analysis dictates that after the spectrum is estimated, its values are used to compute statistics that summarize its shape, e.g. the first four spectral moments, then it is not a foregone conclusion that the two spectral estimators will deliver different results, just by virtue of their having different bias properties.

3.1.2 Variance of the windowed periodogram

While the use of a data window such as the Hamming window can reduce the bias of each ordinate of the periodogram, the ordinates of the Hamming-windowed periodogram

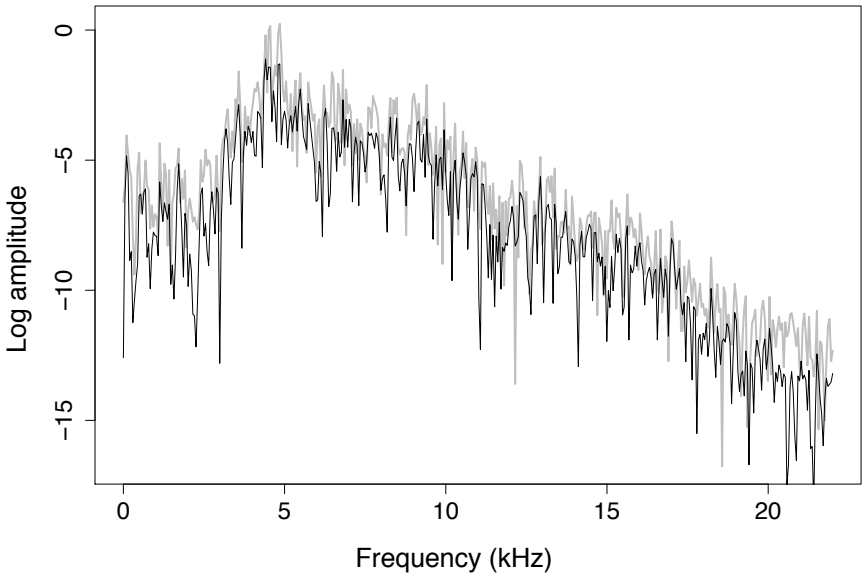

Figure 4: A comparison of a Hamming-window periodogram estimate (thin black line) and a rectangular-window periodogram estimate (thick gray line). Both estimates were computed from the center 20 ms of the token of /s/ shown in Figure 1.

I_{hX} are still prone to having a large MSE because of their large variance. The following theorem, based on Shumway and Stoffer (2006, p. 193, Property P4.2) and stated without proof, establishes the asymptotic distribution of each $I_{hX}(\omega_j)$, from which it is possible to investigate the variance of the estimator's ordinates.

Theorem 3.11 (Distribution of the windowed periodogram ordinates)**.** *If ω_j, $j = 0, 1, \ldots, n-1$, are distinct Fourier frequencies such that $f_X(\omega_j) \neq 0$, for all j, and if for each ω_j, $\{j_n\}$ is a sequence of integers such that $j_n/n \longrightarrow \omega_j$ as $n \longrightarrow \infty$, then as $n \longrightarrow \infty$,*

$$I_{hX}(j_n/n) \xrightarrow{d} \frac{f_X(\omega_j)}{2}\chi^2_2, \tag{37}$$

where $\xrightarrow{d}$ denotes convergence in distribution.

Hence, the variance of each ordinate of a Hamming-windowed periodogram is approximately

$$\begin{aligned} \operatorname{Var}\left[I_{hX}(\omega_j)\right] &\approx \operatorname{Var}\left[\frac{f_X(\omega_j)}{2}\chi^2_2\right] && (38)\\ &= \left(\frac{f_X(\omega_j)}{2}\right)^2 \operatorname{Var}[\chi^2_2] && (39)\\ &= f_X(\omega_j)^2. && (40) \end{aligned}$$

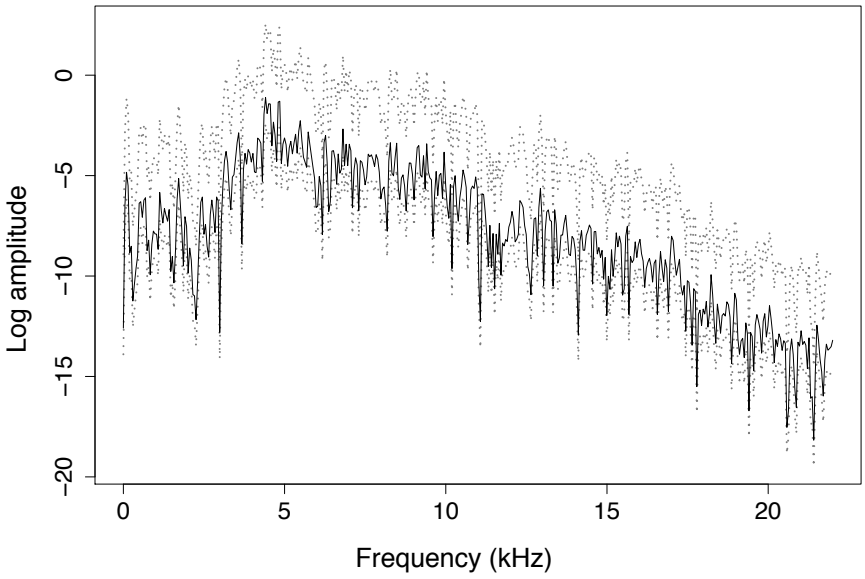

Figure 5: The Hamming-window periodogram estimate (black line) redrawn from Figure 4 plotted with the upper and lower bounds (gray dotted line) of a 95% confidence interval for each ordinate.

The asymptotic distribution of $I_{hX}(\omega_j)$ can also be used to approximate a confidence interval for $f_X(\omega_j)$ with confidence level $(1-\alpha)$. For a given α such that $0 < \alpha < 1$, under the asymptotic distribution of $I_{hX}(\omega_j)$ in (37), there is $(1-\alpha)$ probability that $I_{hX}(\omega_j)$ falls within the interval

$$\frac{f_X(\omega_j)}{2}\chi^2_2(\alpha/2) \leq I_{hX}(\omega_j) \leq \frac{f_X(\omega_j)}{2}\chi^2_2(1-\alpha/2),$$

where $\chi^2_2(\alpha)$, which is referred to as the *lower* α *probability tail*, is the number that satisfies $\mathbb{P}\left(\chi^2_2 < \chi^2_2(\alpha)\right) = \alpha$. Rearranging the terms in the above inequality yields a $100(1-\alpha)\%$ confidence interval for $f_X(\omega_j)$:

$$\frac{2I_{hX}(\omega_j)}{\chi^2_2(1-\alpha/2)} \leq f_X(\omega_j) \leq \frac{2I_{hX}(\omega_j)}{\chi^2_2(\alpha/2)}. \tag{41}$$

These confidence intervals can be visualized by plotting their upper and lower bounds against frequency. Figure 5 shows the Hamming-windowed periodogram estimate from Figure 4 along with the upper and lower bounds of a 95% confidence interval for each ordinate.

Furthermore, the form of the inequality in (41) suggests how the size of each confidence interval can be reduced. Specifically, the size of each confidence interval is directly related to the distance between the lower $(1-\alpha/2)$ and the lower $\alpha/2$ probability tails for the chi-squared distribution with two degrees of freedom. If it were possible to find a spectral estimator whose ordinates had an asymptotic distribution that depended on a distribution

δ whose variance was less than that of χ^2_2, then the distance between the lower $(1-\alpha/2)$ and the lower $\alpha/2$ probability tails of δ would be less than that for χ^2_2, and the size of the confidence interval for each ordinate of the spectral estimator would decrease as well. The desire for such a reduced-variance estimator motivates the introduction of the multitaper spectrum.

3.2 The multitaper spectrum

The multitaper spectrum was introduced by Thomson (1982), and the method of its calculation is simple enough: K copies of a sample $X_1, X_2, \ldots, X_n$ of a stationary process are weighted by K different data windows $\{w_{k,t}\}$. Then, for each windowed realization $\{w_{k,t}x_t\}$, its *eigenspectrum* S_k is found by computing its power spectrum. Finally, the K eigenspectra are averaged to produce the multitaper spectrum $M_X^{(K)}$.

It is also easy to get a sense for why this method would yield a spectral estimator, the variance of whose ordinates is less than that of the Hamming-windowed periodogram. If for a fixed Fourier frequency ω_j, the K ordinates $\{S_k(\omega_j)\}$ all have equal variance and are pairwise uncorrelated, then the ordinate of the multitaper spectrum at that frequency, $M_X^{(K)}(\omega_j)$, will have variance that is $1/K$ the size of the variance of $S_k(\omega_j)$. The aim is therefore to find data windows that will yield uncorrelated eigenspectra whose ordinates each have reasonable variance.

Data windows that satisfy these conditions are found in the family of discrete prolate spheroidal (DPS) sequences (Slepian and Pollak, 1961; Landau and Pollak, 1961, 1962; Slepian, 1964). These sequences were originally discovered as a solution to the spectral concentration problem, which asks whether it is possible to find a sequence of finite duration whose spectrum contains the maximal proportion of its energy in a fixed frequency band. To state the problem more concretely, the Fourier transform of a finite sequence is introduced (Beerends et al., 2003, § 18.5).

Definition 3.12 (Fourier transform of finite sequence)**.** If $x_1, x_2, \ldots, x_n$ is a finite sequence of real numbers, then its Fourier transform $\mathcal{X}$ is defined by

$$\mathcal{X}(\omega) = \sum_{k=1}^{n} x_k e^{-2\pi i \omega k}, \qquad -1/2 \leq \omega \leq 1/2. \tag{42}$$

If $x_1, x_2, \ldots, x_n$ is a sequence of length n, whose Fourier transform is $\mathcal{X}$, and W is a frequency such that $0 < W < 1/2$, then the *spectral concentration* of $\mathcal{X}$ in the band $[-W, W]$, denoted by $\lambda(n, W)$ is defined as

$$\lambda(n, W) = \frac{\int_{-W}^{W} |\mathcal{X}(\omega)|^2 \, \mathrm{d}\omega}{\int_{-\infty}^{\infty} |\mathcal{X}(\omega)|^2 . \, \mathrm{d}\omega} \tag{43}$$

The spectral concentration problem asks whether, given parameters n and W as above, it is possible to find the sequence that maximizes $\lambda(n, W)$. It turns out that the answer to this question is positive (Percival and Walden, 1993, Ch. 3 & 8). Moreover, it is possible to rank the sequences of length n according to their concentration $\lambda(n, W)$. This leads to the definition of a DPS sequence.

Definition 3.13 (DPS sequence)**.** Given fixed parameters n and W to the spectral concentration problem, the *DPS sequence of order k*, denoted by $\{v_t^{(k)}\}$ is the $(k+1)^{\text{th}}$ maximal concentration $\lambda(n, W)$.

So, the sequence that has the greatest energy concentration in the frequency band $[-W, W]$ is the DPS sequence of order 0; the sequence that has the second greatest energy concentration in $[-W, W]$ is the DPS sequence of order 1; and so on.

In order to generate DPS sequences that can be used as data windows for a spectral estimator is it necessary to set the frequency bandwidth parameter W. Conventionally, W is chosen so that the product nW is an integer that satisfies $nW \leq 4$. Futhermore, the choice of W places an upper bound on the number of eigenspectra K that are averaged to compute the multitaper spectrum. In particular, K should satisfy $K \leq 2nW$ (Percival and Walden, 1993, pp. 334-5).

As an illustration, DPS sequences were generated using the `multitaper` package for `R`, with the parameters $n = 883$ (the number of data points in a 20 ms waveform sampled at 44.1 kHz) and $W = 4/883$. For these parameters, the DPS sequences of orders $k = 0$ through $k = 5$ are shown in the top row of Figure 6. The corresponding eigenspectra for the center 20 ms of the /s/ from Figure 1 are shown in the bottom row of that same figure.

It can be shown that the ordinates of each eigenspectrum S_k all have the same asymptotic distribution as the ordinates of the Hamming-window periodogram (Percival and Walden, 1993, p. 343). That is, for all k such that $0 \leq k \leq K$ and j such that $0 \leq j \leq n-1$, the spectral ordinate estimator $S_k(\omega_j)$ converges in distribution to a scaled χ^2_2 random variable.

The DPS sequences are mutually *orthogonal*, in the sense that, for all orders j and k such that $j \neq k$,

$$\sum_{t=1}^{n} v_t^{(j)} \cdot v_t^{(k)} = 0,$$

which ensures that the eigenspectra used in the computation of the multitaper spectrum are pairwise uncorrelated (Percival and Walden, 1993).

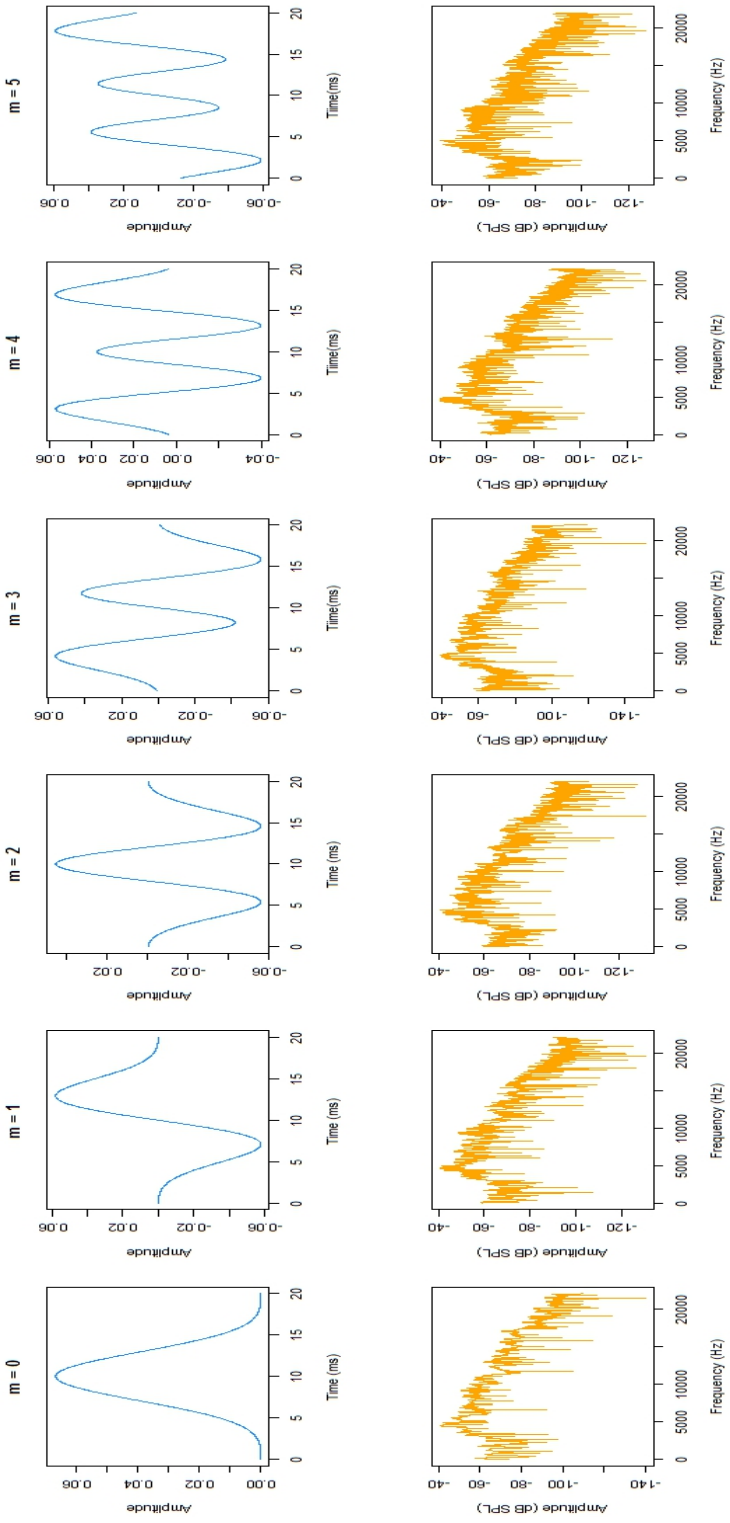

Figure 6: *Top row:* The DPS sequences of order $m = 0$ to $m = 5$, computed using the parameters $n = 883$ and $W = 4/883$. *Bottom row:* The eigenspectrum of the center 20 ms window of the /s/ from Figure 1, each computed using the DPS sequence above as a data window.

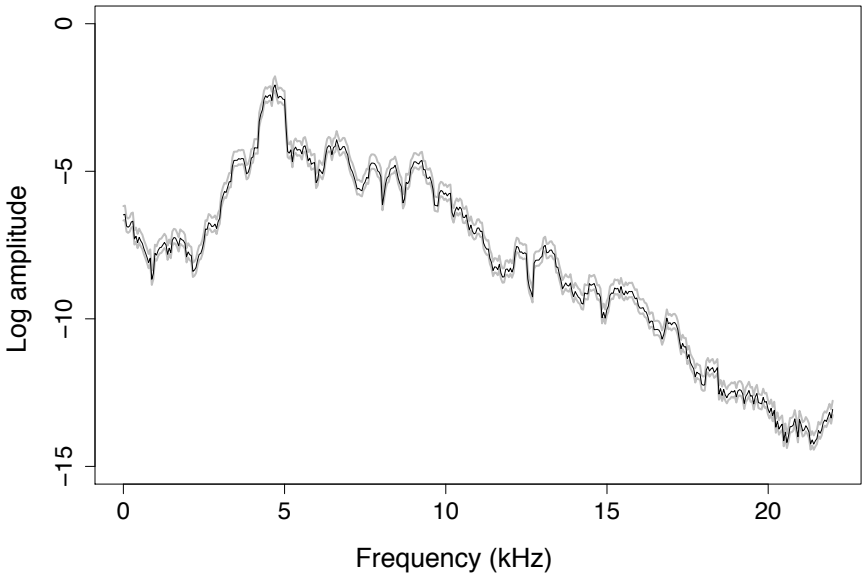

Figure 7: The multitaper spectrum (black line) of the center 20 ms of the /s/ from Figure 1, plotted with the upper and lower bounds (gray lines) of a 95% confidence interval for each ordinate.

The K mutually uncorrelated eigenspectra are averaged pointwise to compute the multitaper spectrum $M_X^{(K)}$; hence, the asymptotic distribution of each ordinate of $M_X^{(K)}$ is a scaled chi-squared with $2K$ degrees of freedom:

$$M_X^{(K)}(\omega_j) = \frac{1}{K}\sum_{k=0}^{K-1} S_k(\omega_j) \xrightarrow{d} \frac{f_X(\omega_j)}{2K}\chi^2_{2K}. \tag{44}$$

Using this distribution to approximate the variance of each ordinate $M_X^{(K)}(\omega_j)$ yields

$$\mathrm{Var}\left[M_X^{(K)}(\omega_j)\right] \;=\; \left(\frac{f_X(\omega_j)}{2K}\right)^2 \mathrm{Var}\left[\chi^2_{2K}\right] \;=\; \frac{f_x(\omega_j)^2}{K}. \tag{45}$$

Comparing (45) to (40), it is obvious that the ordinates of a multitaper spectrum have $1/K$ the variance of the ordinates of a Hamming-window periodogram.

The benefit of this reduced variance is revealed by the size of the confidence intervals for the ordinates of the multitaper spectrum. The confidence interval for each ordinate $M_X^{(K)}(\omega_j)$ is calculated analogously to (41),

$$\frac{2K\, M_X^{(K)}(\omega_j)}{\chi^2_{2K}(1-\alpha/2)} \le f_X(\omega_j) \le \frac{2K\, M_X^{(K)}(\omega_j)}{\chi^2_{2K}(\alpha/2)}. \tag{46}$$

Figure 7 shows the multitaper spectrum of the 20 ms of /s/ plotted with upper and lower bounds of a 95% confidence interval for each ordinate. The size of these confidence intervals is noticeably smaller than those in Figure 5: the mean size of the confidence intervals

for the ordinates of the Hamming-window periodogram estimate is 4.982, while the mean value of those of the multitaper spectrum estimate is 0.486. The correct interpretation of this difference in size is that when estimating an *ordinate* of the spectral density at a given frequency, $f_X(\omega_j)$, it is possible to circumscribe a smaller set of values within which $f_X(\omega_j)$ is likely to fall if the multitaper spectrum is used rather than the Hamming-window periodogram.

This brief introduction to spectral estimation has focused on two spectral estimators that have been used in speech applications: the Hamming-windowed periodogram and the multitaper spectrum. The comparison of these two estimators was carried out primarily in terms of the asymptotic variance of each estimator's ordinates, and it was shown that the multitaper spectrum has a variance that is a fraction of that of the Hamming-windowed periodogram.

4 Conclusion

In this paper, the spectral representation theory for random processes was reviewed, and two methods for estimating the spectrum of a random process were introduced and evaluated. The evaluation of the spectral estimators was carried out in theoretical terms, e.g. by comparing the variance of the asymptotic distribution of each estimator's ordinates. It was shown that the multitaper spectrum is a much "better" estimator than the Hamming-windowed periodogram in the sense that the variance of the former's ordinates is a fraction of that of the latter.

The reader is advised to bear in mind that this notion of "better" is purely theoretical, and in practice a spectral analysis of speech data usually does not end with the estimation of a spectrum, but with the estimation of *properties* of a spectrum, e.g. the peak frequency or one or more of the formants or spectral moments. Therefore, the comparison of the multitaper spectrum and the Hamming-windowed periodogram presented in §3 does not settle the question of which estimator is better-suited to a particular spectral analysis. In fact, it doesn't even address the question since doing so can only be done meaningfully once the details of the analysis are known.

References

Beerends, R. J.; H. G. ter Morsche; J. C. van den Berg; and E. M. van de Vrie. 2003. *Fourier and Laplace transforms*. Cambridge University Press, Cambridge, UK.

Landau, H. J., and H. O. Pollak. 1961. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—ii. *Bell System Technical Journal* 40.65–84.

Landau, H. J., and H. O. Pollak. 1962. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—iii: The dimension of the space of essentially time- and band-limited signals. *Bell System Technical Journal* 41.1295–1336.

Percival, Donald B., and Andrew T. Walden. 1993. *Spectral analysis for physical applications: Multitaper and conventional univariate techniques*. Cambridge University Press, Cambridge, UK.

Shumway, Robert H., and David S. Stoffer. 2006. *Time series analysis and its applications*. Springer Texts in Statistics. Springer, 2^{nd} edition.

Slepian, David. 1964. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—iv: Extensions to many dimensions; generalized prolate spheroidal functions. *Bell System Technical Journal* 43.3009–3058.

Slepian, David, and H. O. Pollak. 1961. Prolate spheroidal wave functions, Fourier analysis, and uncertainty—i. *Bell System Technical Journal* 40.43–64.

Thomson, David J. 1982. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE* 70.1055–1096.

Appendices

The appendices below contain `R` code that can be used for the spectral analysis of speech data—in particular, the computation of Hamming-windowed periodogram and multitaper spectrum estimates.

A `Waveform.r`

```
# Author:       Patrick Reidy
# Affiliations: The Ohio State University
#               Department of Linguistics
#               www.ling.ohio-state.edu
#               Learning To Talk
#               www.learningtotalk.org
# Email:        reidy@ling.ohio-state.edu
# Mail:         24A Oxley Hall
#               1712 Neil Ave.
#               Columbus, OH 43210-1298
# License:      GPL-3


# The Waveform package depends on the Simon Urbanek's
# 'audio' package.
library('audio', quietly=TRUE)
```

```
########################
#  Utility functions  #
########################

`%@%` <- function(...) {
# %@% is a generic function for getting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%.default
  UseMethod('%@%')
}

`%@%.default` <- function(object, attribute) {
# %@%.default is the default method for getting the value of
# an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#                    slot of object.
# Returns:
#   The value of the R object in the attribute slot of
#   object.
  attr(object, attribute)
}

`%@%<-` <- function(...) {
# %@%<- is a generic function for setting the value of an
#   object's attribute(s).
# Methods available in the Waveform package:
#   %@%<-.default
  UseMethod('%@%<-')
}

`%@%<-.default` <- function(object, attribute, value) {
# %@%<-.default is the default method for setting the value
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#                    slot of object.
#       value: Any R object.
# Returns:
#   Nothing.  Instead, the value of object's attribute slot
#   is changed to value.
  `attr<-`(object, attribute, value)
```

```
}

.ConvertUnitNameToMultiplier <- function(unitName) {
# .ConvertUnitNameToMultiplier is a utility function for
# converting the time unit c('second', 'millisecond',
# 'microsecond', 'nanosecond') to proportions of one second.
# .ConverUnitNameToMultiplier implements the following map:
#        'second'   |-->   1
#   'millisecond'   |-->   1000
#   'microsecond'   |-->   1000000
#    'nanosecond'   |-->   1000000000
  if (unitName == 'second') {
    multiplier <- 1
  } else if (unitName == 'millisecond') {
    multiplier <- 1000
  } else if (unitName == 'microsecond') {
    multiplier <- 1000000
  } else if (unitName == 'nanosecond') {
    multiplier <- 1000000000
  }
  return(multiplier)
 }

.FindSampleAtTime <- function(waveform, timeOfSample,
                  timeUnit=(waveform %@% 'timeUnit')) {
# .FindSampleAtTime is a utility function for finding the
# index of the sample that occurs at a given time. Each
# sample point of the Waveform object is conceived of as
# being a half-open interval that is closed on the left and
# open on the right. The time value of each sample point
# is the value of the left boundary.
# Arguments:
#       waveform: A Waveform object.
#   timeOfSample: A numeric specifying the time of the
#                 sample whose index is to be found.
#       timeUnit: A character string specifying the unit of
#                 the timeValue argument.  Legal values:
#                 c('second', 'millisecond', 'microsecond',
#                 'nanosecond').
#                 Default is the unit of time for the time
#                 values of the Waveform object.
# Returns:
#   An integer specifying the index of the sample of the
#   Waveform object that occurs at the time specified by
#   the timeOfSample argument.
```

```
  # Create a vector of the sample times for the Waveform
  # object.
  sample.times <- .ComputeSampleTimes(waveform)

  # Convert the timeOfSample to the same unit as the sample
  # times.
  wave.time.unit <-
   .ConvertUnitNameToMultiplier(waveform %@% 'timeUnit')
  sample.time.unit <-
   .ConvertUnitNameToMultiplier(timeUnit)
  time.of.sample <-
   timeOfSample * (wave.time.unit / sample.time.unit)

  # Find the sample times that are prior or equal to the
  # time.of.sample.
  prior.sample.times <-
   which(sample.times <= time.of.sample)

  # The sample index is the last sample whose time is prior
  # or equal to the time.of.sample; hence, the index is
  # equal to the length of prior.sample.times
  sample.index <- length(prior.sample.times)

  # Return the index of the sample.
  return(sample.index)
}

.ComputeSampleTimes <- function(waveform) {
# .ComputeSampleTimes is a utility function for computing
# the time values of the sampled values of the waveform--
# i.e., the values that are not zeroes padded at the end of
# waveform in the case when waveform has been zero-padded.
# Arguments:
#   waveform: A Waveform object.
# Returns:
#   An integer specifying the number of sampled values in
#   waveform.

  # Find the start time of the waveform.
  start.time <- waveform %@% 'startTime'

  # Find the end time of the waveform.
  end.time <- waveform %@% 'endTime'
```

```
  # Find the number of samples in the waveform.
  sample.n <- waveform %@% 'N'

  # Create a vector of sample times from the start time, end
  # time and number of samples.
  sample.times <-
   seq(from=start.time, to=end.time, length.out=sample.n)

  # Return the vector of sample times.
  return(sample.times)
}

############################
#  Object initialization  #
############################

Waveform <- function(...) {
# Waveform is a generic function for creating a Waveform
# object.
# methods available in Waveform package:
#   Waveform.audioSample
#   Waveform.character
  UseMethod('Waveform')
}

Waveform.audioSample <-
function(audioSample, startTime=0, timeUnit='second') {
# Waveform.audioSample is a method for initializing a
# Waveform object from an audioSample object.  The
# audioSample class is defined in the 'audio' package.
# Arguments:
#   audioSample: An audioSample object.
#     startTime: A numeric that specifies the time of the
#                first sampled value of the audioSample
#                object.  Default is 0.
#      timeUnit: A character string specifying the unit of
#                measurement for startTime and endTime.
#                Legal values: c('second', 'millisecond',
#                   'microsecond', 'nanosecond').
#                Default is 'second'.
# Returns:
#   A Waveform object, which is a numeric vector whose
#   values represent the sampled values of the waveform,
#   augmented with the following attributes:
#        bitRate: The bitrate of the sampled waveform.
```

```
#     sampleRate: The sampling rate of the sampled waveform.
#   samplePeriod: The sampling period of the sampled
#                 waveform.
#              N: The number of samples in the waveform,
#                 excluding those values that are added to
#                 the waveform for zero-padding.
#      startTime: The time at which the first value of the
#                 waveform was sampled.
#        endTime: The time at which the last value of the
#                  waveform was sampled.
#       duration: The duration of the waveform.
#       timeUnit: The unit of measurement for startTime,
#                 endTime, and duration.

  # The audioSample object is a numeric vector that has a
  # 'bits' attribute and a 'rate' attribute for the bit rate and
  # sampling rate, respectively.
  waveform <- as.numeric(audioSample)

  # Set the 'bitRate' attribute of waveform to the value of
  # the 'bits' attribute of audioSample.
  waveform %@% 'bitRate' <- audioSample %@% 'bits'

  # Set the 'sampleRate' attribute of waveform to the value
  # of the 'rate' attribute of audioSample.
  waveform %@% 'sampleRate' <- audioSample %@% 'rate'

  # Set the 'samplePeriod' attribute of waveform.
  waveform %@% 'samplePeriod' <-
    1 / (waveform %@% 'sampleRate')

  # Set the 'N' (number of samples attribute of waveform.
  waveform %@% 'N' <- length(waveform)

  # Set the 'startTime' attribute of waveform from the
  # startTime argument.
  waveform %@% 'startTime' <- startTime

  # Set the 'endTime' attribute of waveform.
  time.lag.from.start <-
   ((waveform %@% 'N') - 1) / (waveform %@% 'sampleRate')
  waveform %@% 'endTime' <-
   time.lag.from.start + (waveform %@% 'startTime')

  # Set the 'duration' attribute of waveform.  The
```

```
  # (waveform %@% 'samplePeriod')  term is added in the
  # calculation below because each sampled point is a
  # treated as a semi-open interval (closed on the left,
  # open on the right) of duration equal to one sample
  # period.
  sampled.time.range <-
   (waveform %@% 'endTime') - (waveform %@% 'startTime')
  waveform %@% 'duration' <-
   sampled.time.range + (waveform %@% 'samplePeriod')

  # Set the 'timeUnit' attribute of waveform from the
  # timeUnit argument.
  waveform %@% 'timeUnit' <- timeUnit

  # Set the class of waveform.
  class(waveform) <- 'Waveform'

  # Return the Waveform object.
  return(waveform)
}

Waveform.character <-
function(waveFilepath, startTime=0, timeUnit='second') {
# Waveform.character is a method for initializing a Waveform
# object from the file path of a .wav file.
# Arguments:
#   waveFilepath: A character string specifying either the
#                 absolute or relative file path of a .wav
#                 file.
#      startTime: A numeric that specifies the time of the
#                 first sampled value of the waveform
#                 pointed to by waveFilepath.  Default is 0.
#       timeUnit: A character string specifying the unit of
#                 measurement for startTime and endTime.
#                 Legal values: c('second', 'millisecond',
#                 'microsecond', 'nanosecond').
#                 Default is 'second'.
# Returns:
#   A Waveform object, which is a numeric vector whose
#   values represent the sampled values of the waveform,
#   augmented with the following attributes:
#        bitRate: The bitrate of the sampled waveform.
#     sampleRate: The sampling rate of the sampled waveform.
#   samplePeriod: The sampling period of the sampled
#                 waveform.
```

```
#      startTime: The time at which the first value of the
#                 waveform was sampled.
#        endTime: The time at which the last value of the
#                 waveform was sampled.
#       duration: The duration of the waveform.
#       timeUnit: The unit of measurement for startTime,
#                 endTime, and duration.

  # Create an audioSample object by loading the wave file
  # pointed to by waveFilepath.
  audioSample <- load.wave(waveFilepath)

  # Dispatch the Waveform.audioSample method.
  Waveform(audioSample, startTime, timeUnit)
}

########################################
#  Methods to R's generic functions  #
########################################

plot.Waveform <- function(waveform, xAxisUnit='millisecond',
   type='l', col='orange',
   xlab=sprintf('Time (%s)', xAxisUnit), ylab='', ...) {
# plot.Waveform is a method for visualizing a Waveform
# object.
# Arguments:
#   waveform: A Waveform object.
#  xAxisUnit: A character string specifying the unit of the
#             time points plotted along the x-axis. Legal
#             values: c('second', 'millisecond',
#            'microsecond', 'nanosecond').
#             Default is 'millisecond'.
#       type: The type of line used to plot the values of
#             waveform. This value is passed to the
#             graphical parameter 'type'.
#        col: The color of the line used to plot the values
#             of waveform. This value is passed to the
#             graphical parameter 'col'.
#       xlab: The label on the x-axis.  This value is passed
#             to the graphical parameter 'xlab'. Default is
#             'Time (<xAxisUnit>)', where <xAxisUnit> is
#             replaced by the value of the xAxisUnit
#             argument.
#       ylab: The label on the y-axis.  This value is passed
#             to the graphical parameter 'ylab'.  Default is to
```

```
#             have no label.
#        ...: Other graphical parameters.
# Returns:
#   A plot of the Waveform object.

  # Make a vector of the time points at which the waveform's
  # samples occur.
  sample.times <- .ComputeSampleTimes(waveform)

  # Convert the unit of sample.times.
  wave.time.unit <-
     .ConvertUnitNameToMultiplier(waveform %@% 'timeUnit')
  x.axis.unit <- .ConvertUnitNameToMultiplier(xAxisUnit)
  sample.times <-
     sample.times * (x.axis.unit / wave.time.unit)

  # Grab just the sampled values of the waveform, excluding
  # the zero-padded values.
  wave.values <- as.numeric(waveform)
  sample.values <- wave.values[1:(waveform %@% 'N')]

  # Plot the Waveform object.
  plot(x=sample.times, y=sample.values,
       type=type, col=col, xlab=xlab, ylab=ylab, ...)
}

print.Waveform <- function(waveform) {
# print.Waveform is a method for reporting the attributes
# and visualizing a Waveform object.
# Arguments:
#   waveform: A Waveform object.
# Returns:
#   A report of the attributes of the waveform are printed
#   to the screen and a plot of the waveform is created.

  # Print the attributes of the Waveform object.
  message(sprintf(
     'Sampling rate:     %.2f', waveform %@% 'sampleRate'))
  message(sprintf(
     'Bit rate:          %d', waveform %@% 'bitRate'))
  message(sprintf(
     'Number of samples: %d', waveform %@% 'N'))
  message(sprintf(
     'Padded to:         %d', length(waveform)))
  message()
```

```
  message(sprintf(
     'Start time:        %f', waveform %@% 'startTime'))
  message(sprintf(
     'End time:          %f', waveform %@% 'endTime'))
  message(sprintf(
     'Duration:          %f', waveform %@% 'duration'))
  message(sprintf(
     'Time unit:         %s', waveform %@% 'timeUnit'))

  # Visualize the Waveform object.
  plot(waveform)
}

################################################
#  New generic functions and Waveform methods  #
################################################

FirstDifference <- function(...) {
  UseMethod('FirstDifference')
}

FirstDifference.Waveform <-
   function(waveform, coefficient=1) {
  # Make a delayed copy of the waveform that is scaled by
  # the coefficient.
  delayed.and.scaled <-
     c(0, waveform[1:(length(waveform)-1)]) * coefficient

  # Subtract the delayed and scaled copy from the waveform.
  preemphed.wave <- waveform - delayed.and.scaled

  # Return the pre-emphasized waveform.
  return(preemphed.wave)
}

TimeSlice <- function(...) {
# TimeSlice is a generic function for slicing a portion of a
# time series-like object according to time values, rather
# than indices.
# Methods available in Waveform package:
#   TimeSlice.Waveform
  UseMethod('TimeSlice')
}

TimeSlice.Waveform <- function(waveform, sliceFrom, sliceTo,
```

```
      centered=FALSE, duration,
      sliceUnit=(waveform %@% 'timeUnit')) {
# TimeSlice.Waveform is a method for slicing a portion of a
# Waveform object according to time values, rather than
# indices.
# Arguments:
#    waveform: A Waveform object.
#   sliceFrom: A numeric specifying the starting time of
#              the sliced portion of the Waveform object.
#     sliceTo: A numeric specifying the end time of the
#              sliced portion of the Waveform object.
#    centered: A boolean value.  If FALSE, then the values
#              of the sliceFrom and the sliceTo arguments
#              are used, and the duration argument is
#              ignored. If TRUE, then the duration argument
#              is used and the center portion of that
#              duration is sliced.
#    duration: A numeric specifying the duration of the
#              portion to be sliced if centered=TRUE.
#   sliceUnit: A character string specifying the unit of
#              time used to specify the sliceFrom and
#              sliceTo times. Legal values: c('second',
#              'millisecond', 'microsecond', 'nanosecond')
#              Default is the same time unit as the Waveform
#              object.
# Returns:
#   The portion of the Waveform object that falls between
#   the sliceFrom and sliceTo times, or a center portion of
#   the Waveform object if centered=TRUE.  If the Waveform
#   object had been zero-padded, then the zero-padding is
#   not appended to the sliced portion of the Waveform
#   object.

  # If the sliced portion is determined by sliceFrom and
  # sliceTo...
  if (! centered) {
    # Find the sample that occurs at the sliceFrom time.
    slice.from.index <- .FindSampleAtTime(waveform,
       timeOfSample=sliceFrom, timeUnit=sliceUnit)

    # Find the sample that occurs at the sliceTo time.
    slice.to.index <- .FindSampleAtTime(waveform,
       timeOfSample=sliceTo, timeUnit=sliceUnit)

    # Slice the waveform using the slice.from and slice.to
```

```
  # indices.
  sliced.wave <- as.numeric(waveform)
  sliced.wave <-
     sliced.wave[slice.from.index:slice.to.index]
} else {
# If the sliced portion is taken from the center of the
# waveform...
  # Compute the time of the midpoint of the waveform, in
  # waveform time units.
  wave.midpoint <- waveform %@% 'startTime' +
     ((waveform %@% 'duration') / 2)

  # Convert wave.midpoint from waveform time units to the
  # time units in which the slice duration is specified.
  wave.unit.factor <- .
     ConvertUnitNameToMultiplier(waveform %@% 'timeUnit')
  slice.unit.factor <-
     .ConvertUnitNameToMultiplier(sliceUnit)
  conversion.factor <-
     slice.unit.factor / wave.unit.factor
  wave.midpoint <- wave.midpoint * conversion.factor

  # Compute the time at the beginning of the sliced
  # portion.
  slice.from.time <- wave.midpoint - (duration / 2)

  # Find the sample that occurs at slice.from.time.
  slice.from.index <- .FindSampleAtTime(waveform,
     timeUnit=sliceUnit, timeOfSample=slice.from.time)

  # Compute the time at the end of the sliced portion.
  slice.to.time <- wave.midpoint + (duration / 2)

  # Find the sample that occurs at slice.to.time.
  slice.to.index <- .FindSampleAtTime(waveform,
     timeOfSample=slice.to.time, timeUnit=sliceUnit)

  # Slice the waveform using the slice.from and
  # slice.to indices.
  sliced.wave <- as.numeric(waveform)
  sliced.wave <-
     sliced.wave[slice.from.index:slice.to.index]
}

# Copy the attributes of waveform over to those of
```

```
  # sliced.waveform.
  attributes(sliced.wave) <- attributes(waveform)

  # Update the 'startTime', 'endTime', 'duration', and 'N'
  # attributes of sliced.wave.
  # First, create a vector of the sample times for the
  # unsliced Waveform object.
  sample.times <- .ComputeSampleTimes(waveform)
  # Second, set the 'startTime' attribute of sliced.wave to
  # the time of the first sliced sample.
  sliced.wave %@% 'startTime' <-
     sample.times[slice.from.index]
  # Third, set the 'endTime' attribute of sliced.wave to the
  # time of the last sliced sample.
  sliced.wave %@% 'endTime' <- sample.times[slice.to.index]
  # Fourth, calculate the duration of sliced.wave.
  sliced.range <- (sliced.wave %@% 'endTime') -
     (sliced.wave %@% 'startTime')
  sliced.wave %@% 'duration' <- sliced.range +
     (sliced.wave %@% 'samplePeriod')
  # Lastly, update the 'N' attribute of sliced.waveform.
  sliced.wave %@% 'N' <- length(sliced.wave)
  # Return the sliced waveform.
  return(sliced.wave)
}

Zeropad <- function(...) {
# Zeropad is a generic function for padding zeroes to the
# end of a time series-like object.
# Methods available in Waveform package:
#   Zeropad.Waveform
  UseMethod('Zeropad')
}

Zeropad.Waveform <-
   function(waveform, lengthOut=(waveform %@% 'sampleRate')) {
# Zeropad.Waveform is a method for padding zeroes to the
# end of a Waveform object.
# Arguments:
#    waveform: A Waveform object.
#   lengthOut: An integer specifying the length of the
#              Waveform object after if has been padded with
#              zeroes.  Default is to pad the waveform to
#              the length equal to its sampling rate.
# Returns:
```

```
#   A Waveform object that is identical to the original
#   Waveform object, but with zeroes added to the end of it.

  # Check that the lengthOut of the padded waveform is
  # greater than the number of sampled values in the
  # waveform.
  if ((waveform %@% 'N') < lengthOut) {
    # If so, pad the waveform.
    # First, make a copy of just the sampled values of the
    # waveform.
    wave.values <- as.numeric(waveform)
    sample.values <- wave.values[1:(waveform %@% 'N')]
    # Second, create a vector of 0's to pad to the end of
    # the sampled values.
    num.zeroes.to.pad <- lengthOut - (waveform %@% 'N')
    zeroes.to.pad <- rep(0, times=num.zeroes.to.pad)
    # Third, pad the zeroes to the sampled values.
    padded.waveform <- c(sample.values, zeroes.to.pad)
    # Lastly, copy the attributes of the Waveform object
    # over to the padded waveform.
    attributes(padded.waveform) <- attributes(waveform)
  } else {
    # If the number of sampled values is greater than the
    # length that the waveform should be padded to, then
    # it cannot be padded.
    padded.waveform <- waveform
    # Print an error message.
    message(
       'You must pad the waveform to a length that is')
    message(
       'number of sampled values in the waveform.')
    message()
    message(sprintf(
       'Number of sampled values: %d', waveform %@% 'N'))
  }

  # Return the padded waveform.
  return(padded.waveform)
}
```

B Tapers.r

```
# Author:       Patrick Reidy
# Affiliations: The Ohio State University
```

```
#               Department of Linguistics
#               www.ling.ohio-state.edu
#               Learning To Talk
#               www.learningtotalk.org
# Email:        reidy@ling.ohio-state.edu
# Mail:         24A Oxley Hall
#               1712 Neil Ave.
#               Columbus, OH 43210-1298
# License:      GPL-3

########################
#  Utility functions  #
########################

`%@%` <- function(...) {
# %@% is a generic function for getting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%.default
  UseMethod('%@%')
}

`%@%.default` <- function(object, attribute) {
# %@%.default is the default method for getting the value
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
# Returns:
#   The value of the R object in the attribute slot of
#   object.
  attr(object, attribute)
}

`%@%<-` <- function(...) {
# %@%<- is a generic function for setting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%<-.default
  UseMethod('%@%<-')
}

`%@%<-.default` <- function(object, attribute, value) {
# %@%<-.default is the default method for setting the value
```

```
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
#       value: Any R object.
# Returns:
#   Nothing. Instead, the value of object's attribute slot
#   is changed to value.
  `attr<-`(object, attribute, value)
}

############################
#  Hamming taper methods  #
############################

Hamming <- function(...) {
  UseMethod('Hamming')
}

Hamming.Waveform <- function(waveform) {
  # Get the number of samples in the Waveform object.
  num.of.samples <- waveform %@% 'N'

  # Generate the sequence of indices for the samples of the
  # Waveform object, that is a sequence of integers from 0
  # to (num.of.samples - 1).
  n.values <- seq(from=0, to=(num.of.samples -1 ))

  # Compute the values of the Hamming window from the
  # sequence of n values.
  hamming.values <- 0.54 -
     (0.46 * cos((2*pi*n.values) / (num.of.samples - 1)))

  # Pad the values of the Hamming window with the same
  # number of 0's that pad the Waveform object.
  zero.pad <-
     rep(0, times=(length(waveform) - num.of.samples))
  hamming.values <- c(hamming.values, zero.pad)

  # Multiply the Waveform object pointwise by the
  # zero-padded Hamming window.
  windowed.wave <- waveform * hamming.values

  # Set the attributes of the windowed waveform.
```

```
  attributes(windowed.wave) <- attributes(waveform)

  # Set an attribute to record how the waveform was
  # windowed.
  windowed.wave %@% 'taper' <- 'Hamming'

  # Return the windowed waveform.
  return(windowed.wave)
}
```

C Periodogram.r

```
# Author:       Patrick Reidy
# Affiliations: The Ohio State University
#               Department of Linguistics
#               www.ling.ohio-state.edu
#               Learning To Talk
#               www.learningtotalk.org
# Email:        reidy@ling.ohio-state.edu
# Mail:         24A Oxley Hall
#               1712 Neil Ave.
#               Columbus, OH 43210-1298
# License:      GPL-3

########################
#  Utility functions  #
########################

`%@%` <- function(...) {
# %@% is a generic function for getting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%.default
  UseMethod('%@%')
}

`%@%.default` <- function(object, attribute) {
# %@%.default is the default method for getting the value of
# an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
# Returns:
```

```
#   The value of the R object in the attribute slot of
#   object.
  attr(object, attribute)
}

`%@%<-` <- function(...) {
# %@%<- is a generic function for setting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%<-.default
  UseMethod('%@%<-')
}

`%@%<-.default` <- function(object, attribute, value) {
# %@%<-.default is the default method for setting the value of
# an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
#       value: Any R object.
# Returns:
#   Nothing. Instead, the value of object's attribute slot
#   is changed to value.
  `attr<-`(object, attribute, value)
}

############################
#  Object initialization  #
############################

Periodogram <- function(...) {
# Periodogram is a generic function for computing the
# ordinate values of the periodogram of a time series-like
# object.
  UseMethod('Periodogram')
}

Periodogram.Waveform <- function(waveform) {
# Periodogram.Waveform is a method for computing the
# ordinate values of the periodogram of a Waveform object.
# Arguments:
#   waveform: A Waveform object.
# Returns:
#   A Periodogram object, comprising the ordinate values of
```

```
#   the periodogram of the Waveform object.  If N is the
#   number of sampled values in the waveform x, then the
#   periodogram I of x is defined by
#     I(w_j) = (1/N) * |d(w_j)|^2,
#   where w_j = j/N is the j^th Fourier frequency (for j =
#   0, ..., N-1) and d(w_j) is the ordinate value of the
#   discrete Fourier transform at w_j, which is defined by
#     d(w_j) = \sum_{t=0}^{N-1} x_t * exp{-2 \pi i w_j t}.
#   A Periodogram object, furthermore, comprises the
#   following attributes:
#         nyquist: The Nyquist frequency of the Waveform
#                  object.
#               N: The number of sampled values in the
#                  Waveform object.
#        binWidth: The width of each frequency bin in the
#                  Periodogram object.
#    fourierFreqs: The hertz values of the Fourier
#                  frequencies.

  # Compute the ordinate values of the periodogram:
  # First, compute the ordinate values of the power spectrum.
  power.spectrum <- abs(fft(waveform))^2
  # Then, scale the power spectrum to get the periodogram.
  periodogram <- (1 / (waveform %@% 'N')) * power.spectrum

  # Keep only the ordinate values that lie on the upper half
  # of the unit circle.  That is, the ordinate values for
  # those frequencies that fall within [0, nyquist).
  nyquist.index <- floor(length(periodogram) / 2)
  periodogram <- periodogram[1:nyquist.index]

  # Set the 'nyquist' attribute of the periodogram, which is
  # equal to half the sampling rate of the Waveform object.
  periodogram %@% 'nyquist' <-
     (waveform %@% 'sampleRate') / 2

  # Set the 'N' attribute of the periodogram, which is equal
  # to the number of sampled values of the Waveform object.
  periodogram %@% 'N' <- waveform %@% 'N'

  # Set the 'binWidth' attribute of the periodogram, which
  # is equal to the sampling rate of the Waveform object,
  # divided by the number of values in the Waveform object
  # (including both sampled and zero-padded values).
  periodogram %@% 'binWidth' <-
```

```
      (waveform %@% 'sampleRate') / length(waveform)

  # Set the 'fourierFreqs' attribute of the periodogram.
  periodogram %@% 'fourierFreqs' <-
      seq(from=0, length.out=nyquist.index,
      by=(periodogram %@% 'binWidth'))

  # Set the class of the periodogram.
  class(periodogram) <-
      c('Periodogram', 'Spectrum', 'numeric')

  # Return the periodogram.
  return(periodogram)
}
```

D Multitaper.r

```
# Author:       Patrick Reidy
# Affiliations: The Ohio State University
#               Department of Linguistics
#               www.ling.ohio-state.edu
#               Learning To Talk
#               www.learningtotalk.org
# Email:        reidy@ling.ohio-state.edu
# Mail:         24A Oxley Hall
#               1712 Neil Ave.
#               Columbus, OH 43210-1298
# License:      GPL-3

# The Multitaper package depends on Karim Rahim's multitaper
# package.
library('multitaper', quietly=TRUE)

########################
#  Utility functions  #
########################

`%@%` <- function(...) {
# %@% is a generic function for getting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%.default
  UseMethod('%@%')
}
```

```
`%@%.default` <- function(object, attribute) {
# %@%.default is the default method for getting the value
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
# Returns:
#   The value of the R object in the attribute slot of
#    object.
  attr(object, attribute)
}

`%@%<-` <- function(...) {
# %@%<- is a generic function for setting the value of an
# object's attribute(s).
# Methods available in the Waveform package:
#   %@%<-.default
  UseMethod('%@%<-')
}

`%@%<-.default` <- function(object, attribute, value) {
# %@%<-.default is the default method for setting the value
# of an object's attribute(s).
# Arguments:
#      object: Any R object.
#   attribute: A character string that names an attribute
#              slot of object.
#       value: Any R object.
# Returns:
#   Nothing.  Instead, the value of object's attribute slot
# is changed to value.
  `attr<-`(object, attribute, value)
}

.ColumnMultiply <- function(numVector, numMatrix) {
# .ColumnMultiply is a utility function for multiplying each
# column of a matrix by a vector.
# Arguments:
#   vect: A numeric vector.
#   matr: A numeric matrix.
# Returns:
#   A matrix that has the same dimensions as matr.  Each
#   column is equal to the corresponding column of matr
```

```
#   multiplied by vect.

  # Multiply each column of matr by vect.
  new.matrix <- apply(numMatrix, 2, `*`, numVector)

  # Return the new matrix.
  return(new.matrix)
}

.NormalizeSum <- function(numVector, normalizeTo=1) {
# .NormalizeSum is a utility function for normalizing the
# values of a numeric vector so that they sum to a
# predetermined number.
# Arguments:
#     numVector: A numeric vector.
#   normalizeTo: A numeric vector of length 1.
# Returns:
#   The numeric vector that results from scaling the
#   elements of numVector so that they sum to sumTo.

  # Determine the scale factor.
  scale.factor <- normalizeTo / sum(numVector)

  # Multiply by the scale factor.
  normalized.vector <- numVector * scale.factor

  # Return the normalized vector.
  return(normalized.vector)
}

############################
#  Object initialization  #
############################

Multitaper <- function(...) {
# Multitaper is a generic function for computing the
# multitaper spectrum of a time series-like object.
  UseMethod('Multitaper')
}

Multitaper.Waveform <- function(waveform, k=(2*nw), nw=4) {
# Multitaper.Waveform is a method for computing the
# multitaper spectrum of a Waveform object.
# Arguments:
#   waveform: A Waveform object.
```

```
#          k: An integer specifying the number of DPSS
#             tapers to use in the computation of the
#             multitaper spectrum.  Default value is
#             k = (2 * nw). Since k and nw are constrained
#             to satisfy k <= 2*nw, the default value is the
#             maximum number of tapers that should be used
#             given a fixed value of nw.
#         nw: An integer specifying the time-bandwidth
#             parameter used to generate the DPSS tapers.
# Returns:
#   The kˆth-order multitaper spectrum of the waveform,
#   computed using DPSS tapers generated using the
#   time-bandwidth parameter nw.  A Multitaper object,
#   furthermore, has the following attributes:
#         nyquist: The Nyquist frequency of the Waveform
#                  object.
#               N: The number of sampled values in the
#                  Waveform object.
#        binWidth: The width of each frequency bin in the
#                  Multitaper object.
#    fourierFreqs: The hertz values of the Fourier
#                  frequencies.
#               k: The number of DPSS tapers (equivalently,
#                  eigenspectra) used in the computation of
#                  the multitaper spectrum.
#              nw: The time-bandwidth parameter used to
#                  generate the DPSS tapers.

  # Generate the DPSS tapers using the dpss function from
  # the multitaper package.  dpss creates a named list whose
  # 'v' element is a matrix, each column of which is a DPSS
  # taper.
  dpss.taper.matrix <-
     dpss(n=(waveform %@% 'N'), k=k, nw=nw)$v

  # Zero-pad the DPSS tapers to the length of the waveform,
  # since the waveform is zero-padded by the difference
  # between length(waveform) and (waveform %@% 'N'):
  # First, determine how many zeroes were padded on the end
  # of the waveform.
  pad.length <- length(waveform) - (waveform %@% 'N')
  # Second, create a matrix of 0's that has pad.length rows
  # and k columns.
  zeropad.matrix <- matrix(data=0, nrow=pad.length, ncol=k)
  # Lastly, row bind the zeropad.matrix to the bottom of the
```

```
# dpss.taper.matrix.
dpss.taper.matrix <- r
   bind(dpss.taper.matrix, zeropad.matrix)

# Make k tapered copies of the sampled waveform by
# windowing it by each DPSS taper, using the
# .ColumnMultiply function.
tapered.wave.matrix <-
   .ColumnMultiply(waveform, dpss.taper.matrix)

# Compute the k eigenspectra of the waveform.  The k^th
# "eigenspectrum" of the waveform is the periodogram of
# the waveform after it has been windowed by the k^th
# DPSS taper.
eigenspectra <- abs(fft(tapered.wave.matrix))^2

# For each eigenspectrum, keep only the ordinate values
# for the frequencies in the [0, nyquist) range.
nyquist.index <- floor(length(waveform) / 2)
nyquist.eigenspectra <- eigenspectra[1:nyquist.index, ]

# Compute the k^th order multitaper spectrum by averaging
# the k eigenspectra, pointwise.
if (k == 1) {
  multitaper <- nyquist.eigenspectra
} else {
  multitaper <- rowMeans(nyquist.eigenspectra)
}

# Set the 'nyquist' attribute of the multitaper spectrum,
# which is equal to half the sampling rate of the
# Waveform object.
multitaper %@% 'nyquist' <-
   (waveform %@% 'sampleRate') / 2

# Set the 'N' attribute of the multitaper spectrum, which
# is equal to the number of sampled values of the Waveform
# object.
multitaper %@% 'N' <- waveform %@% 'N'

# Set the 'binWidth' attribute of the multitaper spectrum,
# which is equal to the sampling rate of the Waveform
# object, divided by the number of values in the Waveform
# object (including both sampled and zero-padded values).
multitaper %@% 'binWidth' <-
```

```
      (waveform %@% 'sampleRate') / length(wavefofrm)

  # Set the 'fourierFreqs' attribute of the multitaper
  # spectrum.
  multitaper %@% 'fourierFreqs' <-
     seq(from=0, length.out=nyquist.index,
     by=(multitaper %@% 'binWidth'))

  # Set the 'k' attribute of the multitaper spectrum.
  multitaper %@% 'k' <- k

  # Set the 'nw' attribute of the multitaper spectrum.
  multitaper %@% 'nw' <- nw

  # Set the class of the multitaper spectrum.
  class(multitaper) <- c('Multitaper', 'Spectrum', 'numeric')

  # Return the multitaper spectrum.
  return(multitaper)
}
```

AN ACOUSTIC ANALYSIS OF VOICING IN AMERICAN ENGLISH DENTAL FRICATIVES[1]

Bridget Smith
The Ohio State University

Abstract

In this study, an acoustic analysis of the dental fricatives as produced by American English speakers from the Buckeye Corpus (Pitt et al. 2006) reveals that the dental fricatives are subject to variation in voicing based on phonetic environment, much more than is usual for a pair of phonemes whose phonological distinction is based on voicing, confirmed by a comparison with the voicing of /f/ and /v/. The results of the study show that voicing (presence or absence of glottal pulses) for /θ/ and /ð/ is not predictable by phoneme in conversational speech, but it is more predictable based on voicing of surrounding sounds.

[1] This acoustic analysis was originally intended to be included in Smith 2009, "Dental fricatives and stops in Germanic." However, the editors and reviewers decided, and rightly so, that there were actually two papers there, so the acoustic analysis was removed and has been referred to as Smith 2007, "The seeds of sound change don't fall far from the tree," unpublished ms. While it has been my intention for some time to replicate the original study, introducing a number of new measurements and methods of analysis, while verifying the original measurements, this paper has sat on a back burner, still unpublished. While I still intend to perform such a reanalysis and submit it to a peer-reviewed journal, under whose scrutiny it will no doubt improve greatly, I submit this early version to the OSU Working Papers so that it may be more easily accessed by those who have expressed interest in it, for those who cannot wait for the next version to be published, and for those who have yet to discover it.

Bridget Smith. An acoustic analysis of voicing in American English dental fricatives. *OSUWPL*, Volume 60, Spring 2013, 117-128.

1. Introduction

The voiceless dental fricative (IPA /θ/) and the voiced dental fricative (IPA /ð/) are relatively understudied in comparison to other sounds. They are perceptually weak and easily confusable with /f/ and /v/. Because the voiced phoneme /ð/ appears word-initially in function words and word-finally in verbs (where the voiceless phoneme /θ/ does not appear), it is difficult to create psycholinguistic tasks in which the salience of the contrast can be measured. And because /θ/ and /ð/ may vary along multiple acoustic dimensions at once, forced choice tasks along a single dimension may not be particularly informative. Categorical replacement of /θ/ and /ð/ by /f/ and /v/ or /t/ and /d/ in certain dialects and sociolects has been studied, such as in AAVE (Wolfram 1970, 1974, among others) or London Cockney English (Wells 1982; Hughes, Trudgill & Watt 2005, among others). Polka, Colantonio, & Sundara (2001) found that English-speaking infants were less able to distinguish between /d/ and /ð/ than between /b/ and /v/, which suggests that the variation in production of the dental fricative which infants are exposed to overlaps to some extent with the alveolar (or dental) stop, so that they are unable to interpret a phonemic pattern until they are much older. A number of studies have looked at acoustic measurements to distinguish place of articulation among various groups of fricatives, and have either avoided /θ/ and /ð/, or were least successful in distinguishing /θ/ from /f/.

Stevens et al. (1992) show that the presence or absence of phonation is the acoustic parameter that best distinguishes between voiced and voiceless fricatives, and they are by no means alone in this judgment. See, e.g., Pirello, Blumstein, & Kurowski (1997). Denes (1955), however, demonstrated earlier that duration of a word-final fricative, and comparatively, the duration of the preceding vowel, could be manipulated to give the impression of voicing for longer vowels and shorter fricatives, and of voicelessness for shorter vowels and longer fricatives. Raphael (1972) confirmed these findings, and noted that "when the voicing characteristic is cued by vowel duration, perception is continuous rather than categorical" (1296). Pirello et al. (1997) say also that the production aspect of voicing is itself continuous, in that "the feature voicing in fricatives is manifested in a continuous way and as such cannot be characterized in terms of a binary distinction relating to the presence or absence of glottal excitation" (3754). Due to the complex nature of producing a voiced fricative, either voicing or frication may be lost during production: too little supraglottal pressure and frication fails, but if the supraglottal pressure is not sufficiently less than the subglottal pressure, voicing will fail. In addition, coarticulation frequently results in gestural overlap, with voicing (or the lack thereof) spreading from neighboring segments. It is for these reasons that a binary presence or absence of glottal pulses is insufficient to discriminate between voiced and voiceless fricatives. The amount of voicing overlap and the issue of duration is greatest for medial and final fricatives, but Pirello et al. (1997) achieved 93% accuracy classifying word-initial /s/ and /z/, and /f/ and /v/ from read speech. Using Stevens et al.'s (1992) rubric, in which the amplitude of the first harmonic of the fricative was compared to that of the following vowel, the fricatives were categorized as voiced and voiceless. The voiceless label was assigned to tokens with 10 dB or greater difference in amplitude. They argued that presence of glottal excitation present in at least 30 ms of either the beginning or the

end of the fricative was enough to correctly distinguish voiced from voiceless fricatives; although, they did not examine /θ/ and /ð/.

These previous studies, and others, have relied upon lab-produced read speech. While this allows researchers to exert some measure of control over variation, and creates tokens that can be easily compared across speakers, it does not give us a real picture of what these fricatives look like in conversational speech, and does not give an accurate picture of how these sounds might be discriminated by language users. Accumulated anecdotal observations provided the questions for this study: How strong is the voicing distinction between /ð/ and /θ/ in conversational speech? And is this distinction based on phonation, or do duration, intensity, or even manner of articulation play a greater role in conversational speech? When does voicing occur? What phonetic factors may be related to voicing? Does the phonological description of 'phoneme' match up with the phonetic realizations of /ð/ and /θ/? A parallel examination of /f/ and /v/ was conducted to find out what measures might be significant in distinguishing the voiced from the voiceless segments, and to provide a control group for comparison.

2. History

Both the voiced and voiceless dental fricatives are represented orthographically in English by the digraph <th>. /θ/ usually occurs word-initially or word-finally, and can occur medially in loanwords and certain compounds. /ð/ occurs word-initially only in function words such as articles and demonstratives, rarely in word-final position in certain derivational words such as *bathe* or *teethe*, and occurs in medial position in a greater number of words. There are a few minimal pairs, such as *thigh* and *thy, either* and *ether,* and *teeth* and *teethe*, and some near-minimal pairs such as *breath* and *breathe.* Despite the existence of at least one minimal pair in all positions, the contrast between these phonemes carries little, if any, functional load. The minimal pairs that exist cannot be used in the same position in a sentence, belonging generally to different classes of words.

This distribution is easily explained through the historical development of these sounds. In Old English, the *thorn* <þ> and *edh* <ð> characters interchangeably represented both the voiced and voiceless variant, which were at that time in complementary distribution. It is generally assumed that thorn or edh in initial and final position was voiceless, while between voiced sounds, it was voiced. /f/ and /v/ (and /s/ and /z/) had a similar distribution. Early on in Middle English, /f/ and /v/ (and /s/ and /z/) became phonemic, due to a confluence of factors, not least of which was the introduction of large numbers of loanwords containing these sounds in contrastive positions. Late during the Middle English period, it was noticed that there were two <th> variants, presumably a voiced and a voiceless phoneme (Bullokar 1580). Function words such as *the, that, this, then*, etc., are assumed to have begun with a voiceless dental fricative in Old English, but their Modern English counterparts have become voiced. Because they are often unstressed and not discrete from adjacent words, they are more likely to assimilate to surrounding voiced sounds. Another possible contributing factor is that the high frequency of these function words may have allowed a large amount of variation,

which became generalized as a voicing contrast. Note that the phonologization of these sounds occurred after the paradigm leveling that reduced the number of different forms of these function words. For example, the definite article *the* was inflected for case, gender, and number in Old English, yielding approximately 12 distinct forms of this word. The increased frequency of single forms of certain types of words may have created the situation that allowed reduction and variation of these high frequency words that now carry much less grammatical information. Word-final /ð/ also appeared in late Middle or Early Modern English, with the loss of verb endings stranding the medially-voiced fricative at the end of the verb. The greatest number of instances of /θ/ that occur outside of the original conditioning environment are in more recent loanwords such as *author, arithmetic,* and *arthritis*, and in forms that have undergone some kind of analogy or reanalysis, such as *Arthur* or *anthem*. /v/ and /f/ have followed a similar trajectory, with fossilization of the original conditioning environment in many words, but with a much greater number of minimal pairs from borrowing, though the contrast is arguably less than that of /s/ and /z/. Therefore, it stands to reason that /v/ and /f/ are the closest point of comparison to /ð/ and /θ/.

3. Methods

Eight talkers (four men, four women) were selected from the Buckeye Corpus (Pitt et al. 2006), which is a body of 40 sociolinguistic interviews with Ohio residents. After subsequent analysis, one of the male talkers' data were excluded from this analysis because of speech differences possibly resulting from a head injury. From approximately 15 minutes of conversational speech per talker, around 400 /ð/ and /θ/ and 300 /v/ and /f/ tokens were measured altogether. Between 15 and 25 /θ/ tokens were measured from each speaker, 25-45 /ð/ tokens, 11-30 /f/ tokens, and 10-35 /v/ tokens. Additional tokens were marked as assimilated or deleted, but not included in the measurements.

In Praat (Boersma & Weenink 2007), an acoustic analysis software, intervals were measured for the duration of frication of /θ/, /ð/, /f/, and /v/. Average intensity of these periods was also measured. Because of the variability of conversational speech, and the greater degree of coarticulation, no single uniform cue existed for determining the beginning or end of the fricative. Obvious frication was used as a marker, where it was present, generally in voiceless tokens. In voiced fricatives, either frication or reduced amplitude of formants as compared to surrounding vowels was a reliable marker. In some cases, /ð/ could be perceived in the speech signal, but not identified by characteristics in the waveform or spectrogram. In these and other cases in which the sound was completely unable to be differentiated from surrounding sounds, it was marked as assimilated or deleted. The beginning and ending of periods of voicing within the fricative were also measured. Intervals were marked as voiced if periodicity in the waveform and/or presence of a voice bar in the spectrogram indicated regular glottal pulses. To minimize confusion with the phonological feature [+voice], these are described as having *voice bar*. A measure of voicing was then created by taking the percentage of the duration of the fricative in which a voice bar could be found. The immediately preceding and following segments adjacent to each fricative, and whether or not they were voiced, were marked and noted. Neighboring segments were only classified

as voiceless if there was no periodicity in the waveform for more than 50% of the segment. A neighboring pause also counted as a voiceless segment because the vocal folds were not vibrating.

A subset of four talkers was selected for an additional analysis examining manner of articulation of the dental fricatives as well. The author examined the waveform and spectrogram, and listened to each token to arrive at the manner judgments. A description of characteristics used to identify each manner is given in the appendix. A second trained transcriber also gave manner judgments for each segment. For the fewer than 5% of the judgments that did not match between the two transcribers, these were then judged in tandem and an appropriate label agreed on. The majority of cases were approximants or flaps that could not be decided on, and without any concrete way of separating these two, the categories were combined into one.

4. Results

The results of the manner judgments are listed in Table 1. /ð/ was realized as a voiced fricative only 20.3% of the time. 23.4% of the time it was realized as a nasal, 18.4% as a stop, and 15.4% as an approximant or flap. /θ/, on the other hand was realized as a voiceless fricative 55.1% of the time, but was also realized as a stop 15.9% of the time, and as a voiced fricative 10.1% of the time. Because of the lack of contrast in manner at the interdental/dental place of articulation, /ð/ and /θ/ are free to vary in this way, though there is much more variation than one might expect.

Table 1: Manner judgments of dental fricatives

Manner of articulation	**edh /ð/**	**theta /θ/**
affricate	0.013	0.029
approximant/flap	0.152	0.043
deletion/vowel	0.095	0.015
voiced fricative	0.203	0.101
voiceless fricative	0.088	0.551
fricative+approximant	0.006	0.0
fricative+stop	0.019	0.058
lateral	0.006	0.0
nasal	0.234	0.015
stop	0.184	0.159
stop+lateral	0.0	0.029

Tokens which were completely assimilated or deleted were not included in the acoustic measurements, because they could not be segmented away from the following sounds. An additional 2 /v/ tokens and 5 /θ/ tokens were removed as outliers, having durations greater than 3 standard deviations from the mean, and were greatly lengthened while the interviewee was thinking of what he/she wanted to say next.

Measurements of absolute duration, as in Figure 1, reveal a great amount of overlap, which is unsurprising for conversational speech. What is notable, however, is that /ð/ and /θ/ pattern very similarly, while /f/ and /v/ show a bimodal distribution, with /f/ having a shorter duration (median 11 ms) than /v/ (median 35 ms), which is entirely the opposite pattern from we would expect. /ð/ and /θ/, while having a similar distribution to each other, also have the majority of tokens clustered around a very short duration, with half the tokens of both being less than 15 ms. These tokens are initial, medial, and final, and in a variety of mono- and polysyllabic words with varying stress. Many neighboring segments were not vowels, and many vowels were reduced to the point of deletion. Calculating the fricative duration relative to surrounding vowel duration did not make sense in this case. Measures of intensity posed the same challenges due to the varied nature of conversational speech. As can be seen in Figure 2, the distribution of average intensity measured over the duration of each fricative is very similar for /ð/ and /θ/ as well as /f/ and /v/, though /f/ has a slightly lower average intensity than /v/.

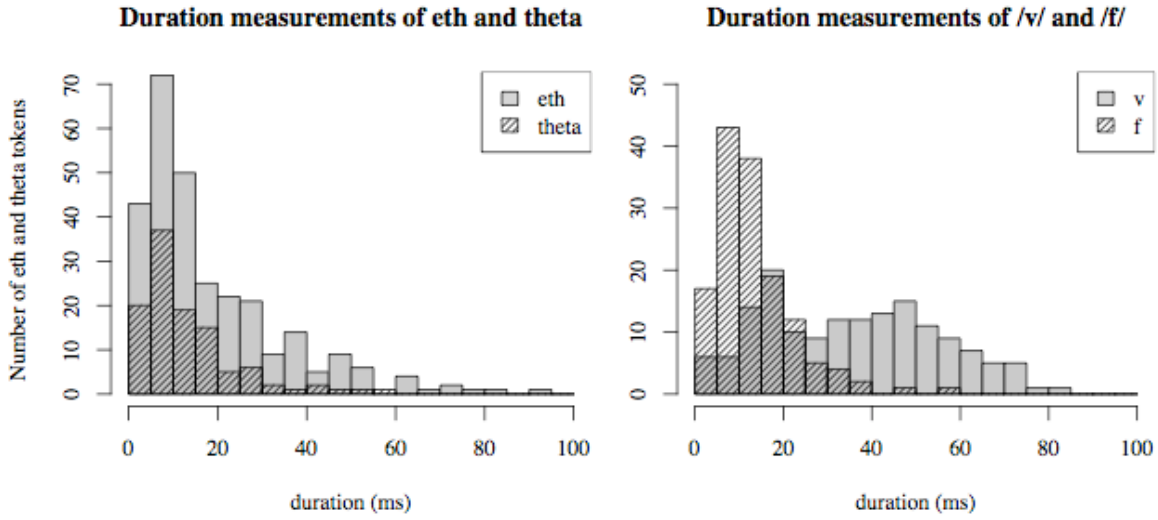

Figure 1: Duration of /ð/ and /θ/ (histogram on the left) and of /f/ and /v/ (on the right).

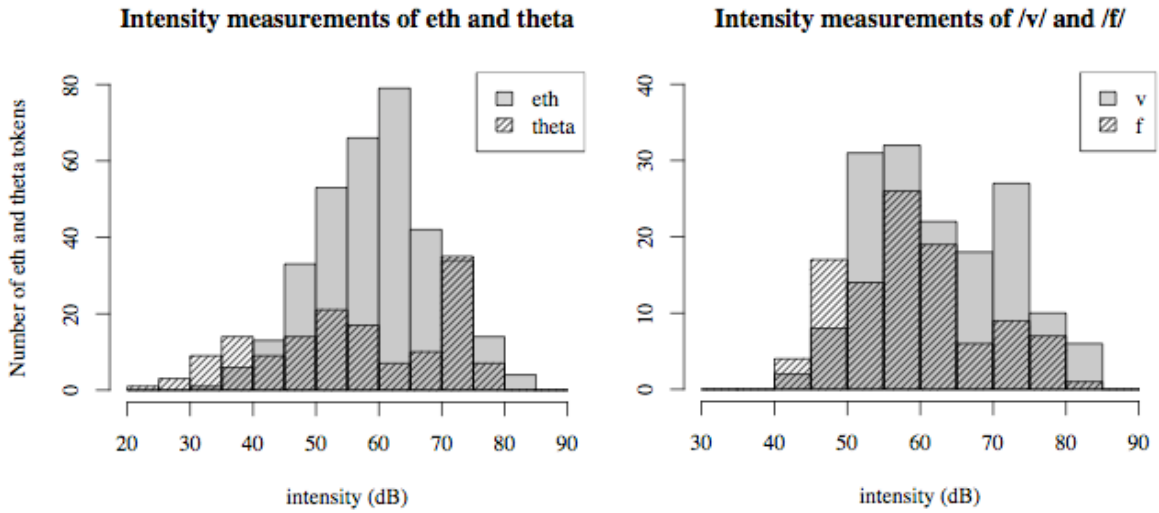

Figure 2: Average intensity of /ð/ and /θ/ (histogram on the left) and of /f/ and /v/ (right).

Because the overall duration measurements were more varied, and generally shorter than those in either Pirello et al. (1997) or Stevens et al. (1992), looking for 30 ms of voicing at either end of the fricative did not make sense. More than half of the fricatives were shorter than 30 ms overall. Because the tokens were not taken from CV sequences, but rather had very different environments, the measurement from Pirello et al. (1997) and Stevens et al. (1992), comparing the intensity of the first harmonic between the fricative and following vowel was not practical. Instead, the appearance of voice bar (periodicity of the waveform) was measured for each fricative interval, and taken as a percentage of the total duration of the fricative, as in Figure 3. The distribution for /f/ and /v/ appears bimodal, although there is a large amount of overlap. The majority of /f/ tokens are clustered below 20% voiced, while the majority of /v/ tokens are greater than 90% voiced. Again, /ð/ and /θ/ do not pattern exactly the same way as /f/ and /v/. There are two distributions, but not separated by phoneme. There is one distribution composed of both /ð/ and /θ/ tokens, clustered around 10-30% voiced, and another substantial cluster of 90-100% voiced tokens, composed primarily of /ð/ tokens, though including a few /θ/ tokens as well.

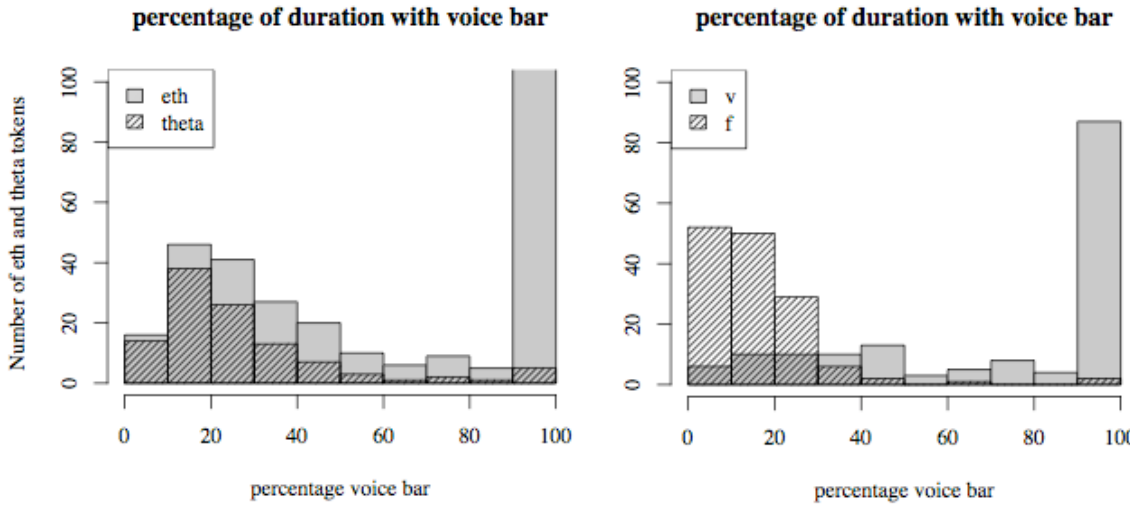

Figure 3: Percentage of the duration of the fricative that contained voice bar, with /θ/ and /ð/ on the left, and /f/ and /v/ on the right.

Yet somehow listeners are able to understand conversational speech. Because /ð/ and /θ/, and to a lesser extent, /f/ and /v/, maintain a similar pattern to their historical distribution, we might expect the voicing to pattern the same regardless of whether we separate the results by phoneme or by environment (that is, whether there is an adjacent voiceless sound, or if the segment is surrounded by voiced sounds). However, when we plot the same voice bar measurements as in Figure 3 by environment, as in Figure 4, a new pattern emerges. The dental fricatives develop a bimodal distribution, with predominantly voiceless sounds clustered on the left and predominantly voiced sounds on the right, distinguished not by phoneme, but by whether they were surrounded by voiced sounds or had a voiceless sound (including a pause) either preceding or following. The distribution of labio-dental fricatives shows a similar though somewhat murkier pattern, where this analysis appears to create more overlap, rather than reducing it.

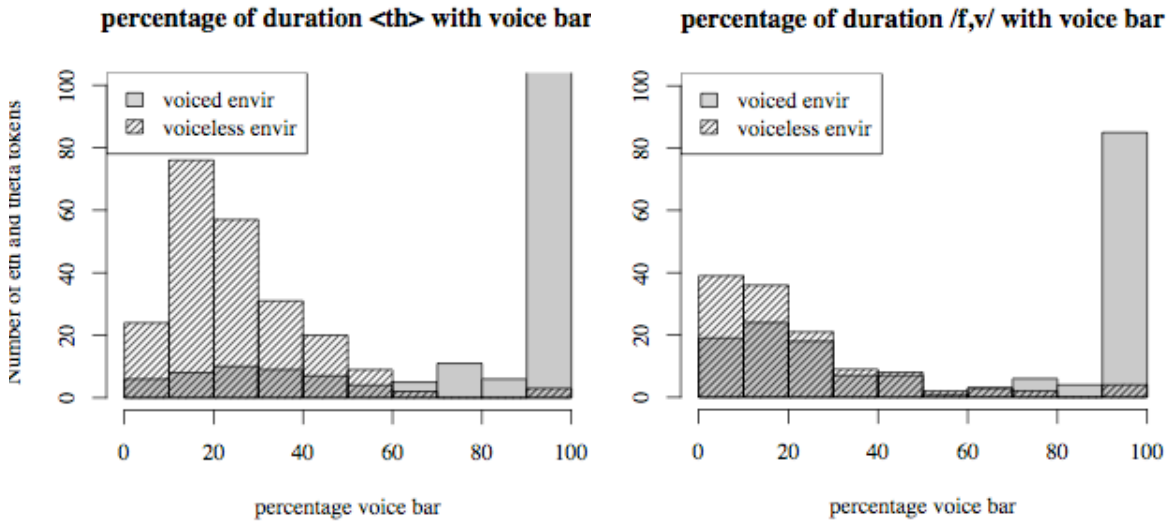

Figure 4: Percentage of the duration of the fricative that contained voice bar, with /θ/ and /ð/ combined on the left, separated by whether or not an adjacent segment was voiceless, and /f/ and /v/ combined on the right, also separated by voicing of adjacent segment.

Because the categories of phoneme and of environmental voicing largely overlap for these sounds, a partial correlation statistic was run on each of these data sets in order to determine how much variation was accounted for by phoneme, excluding that accounted for by environmental condition, and how much was accounted for by environmental condition, excluding that accounted for by phoneme, as in Table 2. Table 3 shows, by way of comparison, the results of a normal correlation of percent voiced by phoneme and by environment, separately.

Table 2: Partial correlation comparing variation accounted for by phoneme and environment, "partialing out" the variation accounted for by the other factor (environment and phoneme, respectively).

	r of voicing by ***phoneme***, with environment partialed out	**r** of voicing by ***environment***, with phoneme partialed out
dental fricative	0.4494	0.7661
labio-dental fricative	0.7186	0.4632

Table 3: Correlation of voicing and environment, and of voicing and phoneme.

	r of voicing by ***phoneme***, regardless of environment	**r** of voicing by ***environment***, regardless of phoneme
dental fricative	0.3666	0.7428
labio-dental fricative	0.7403	0.5157

The results of the correlation and partial regression show that phoneme category better accounts for the percentage of the duration that has voice bar in labio-dental fricatives, such that /v/ has generally more voicing, and /f/ less. But environment, that is, whether surrounding sounds are voiced, does a better job of accounting for percentage of voicing in dental fricatives, regardless of phoneme.

5. Discussion

The results of this investigation seem to indicate that the voicing contrast between the dental fricatives does not hold up well in conversational speech, even less well than /f/ and /v/, which are also highly variable. Duration and intensity are also not reliable measures of phonemic voicing in conversational speech. In fact, the very environments which conditioned voicing of the Old English dental fricative are more or less the same environments that predict voice bar in the modern dental fricatives. Even the "voiced" phonemes in function words can be produced as voiceless, although these are the same words that we assume were at least partially responsible for the phonologization of the voicing contrast. These results call into question whether the distinguishing feature between the two phonemes is actually voicing, or if there are some environments in which the contrast is neutralized. Because there is little to no competition in the interdental place of articulation, manner is free to vary, and indeed only 55% of the tokens of /θ/ and 29% of /ð/ that were analyzed in this experiment were realized as canonical fricatives; therefore, manner is also not a reliable indicator.

One thing that can be said for certain is that the variation of /ð/ occupies a much larger acoustic space, and encompasses that of /θ/, as illustrated in figures 1-3 for duration, intensity, and voicing, but also in place and manner, especially as it assimilates more readily to surrounding sounds, as in the greater number of tokens that were assimilated or deleted in this study. This raises some interesting questions. If the canonical forms are reported to be fairly confusable, but the conversational forms are even more variable and overlapping, how is it that we are able to distinguish them in speech? Or, because of the very low functional load carried by this contrast, do we even need to distinguish them? If voicing is predictable from environment, would it not be likely that listeners would have difficulty reconstructing an underlying form? Future studies are needed to explore perceptual aspects of and contrast or lack thereof in the dental fricatives.

References

Boersma, Paul & David Weenink. 2007. Praat: Doing phonetics by computer (Version 4.5.14) [Computer program]. Retrieved February 2007, from [http://www.praat.org/].

Bybee, Joan, & Paul Hopper. 2001. Editors. *Frequency and the emergence of linguistic structure.* Amsterdam: John Benjamins.

Dalen, Arnold. 2002. "Sources of written and oral languages in the 19th century". In *The Nordic Languages: An international handbook of the history of the North Germanic languages.* Vol 2. Ed. by Oskar Bandle, Kurt Braunmuller, Ernst Hakon Jahr, Allan Karker, Hans-Peter Naumann, & Ulf Teleman. Berlin: DeGruyter. 1406-1418.

Hughes, Arthur, Peter Trudgill, & Dominic Watt. 2005. *English accents and dialects: An introduction to social and regional varieties in the British Isles*. 4th ed. London: Trans-Atlantic Publications, Inc.

Maddieson, Ian, & Kristin Precoda. 1990. UPSID-PC *The UCLA Phonological Segment Inventory Database.* (Data on the phonological systems of 451 languages, with programs to access it.) Accessed from [http://www.linguistics.ucla.edu/faciliti /sales/software.htm]

Pirello, Karen, Sheila E. Blumstein, & Kathleen Kurowski. 1997. "The Characteristics of Voicing in Syllable-Initial Fricatives in American English". *Journal of the Acoustical Society of America* 101.3754-3765.

Pitt, Mark A., Laura Dilley, Keith Johnson, Scott Kiesling, William Raymond, Elizabeth Hume, & Eric Fosler-Lussier, comps. 2006. *Buckeye Corpus of Conversational Speech*. (1st release) [www.buckeyecorpus.osu.edu] Columbus, Ohio: Department of Psychology, Ohio State University (Distributor).

Polka, Linda, Connie Colantonio, & Megha Sundara. 2001. "A cross-language comparison of /d/ - /ð/ perception: Evidence for a new developmental pattern." *Journal of the Acoustical Society of America.* 109(5):2190-2201.

Raphael, Lawrence J. 1972. "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English." *Journal of the Acoustical Society of America.* 51:1296-1303.

Smith, Bridget J. 2007. "The seeds of sound change don't fall far from the tree". Unpublished ms. presented at *OSU Colloquium Fest*, Nov. 9, 2007.

Smith, Bridget J. 2009. "Dental fricatives and stops in Germanic." In Dufresne, Dupoux, & Vocaj (eds.), *Historical Linguistics 2007. CILT 308*. 19-36. Amsterdam/Philadelphia: John Benjamins.

Stevens, Kenneth N., Sheila E. Blumstein, Laura Glicksman, Martha Burton, & Kathleen Kurowski. 1992. "Acoustic and Perceptual Characteristics of Voicing in Fricatives and Fricative Clusters". *Journal of the Acoustical Society of America.* 91: 2979-3000.

Wells, John C. 1982. *Accents of English 2: The British Isles*. Cambridge University Press. 1982.

Wolfram, Walt. 1970. "Some illustrative features of Black English." Workshop on Language Differences. Coral Gables, FL, February 1970

Wolfram, Walt. 1974. *Sociolinguistic aspects of assimilation: Puerto Rican English in New York City.* Arlington, VA: Center for Applied Linguistics.

Appendix: Characteristics of manner of articulation

Criteria used for manner judgments presented in Table 1.

(A) Voiceless fricative: Characterized by aperiodic noise, sometimes forming a diamond or triangle shape in the waveform as intensity increases over the course of the fricative. Because /ð/ and /θ/ are very quiet, the aperiodic noise may maintain a very low intensity throughout. It sounds like a traditional <th> sound, as in *thing*. There may be one or more quiet bursts, but if the frication leading up to and following the burst are of a similar character, and the burst is quiet, and the segment sounds like a fricative, it is labeled as a fricative rather than a stop. Bursts may occur in fricatives as the oral air pressure builds up behind the constriction and is released. This can happen if the lungs are sending out a little too much air to fit through the narrowed constriction. Large bursts occur (and may be labeled as a stop) if the constriction is too narrow and the air pressure too high, so that the air gets really backed up and the constriction acts as a (leaky) closure. See section (C) on stops.

(B) Voiced fricative: Sounds like <th> in the word *then* or *breathe*. There should be frication in the higher frequencies, visible as aperiodic noise in the waveform, as well as regular periodicity indicating voicing for most of the duration. Sometimes the frication may be quiet, and only barely distinguishable from background noise, but there should be at least some fuzziness (aperiodic noise) in each period. Though there will be regular voicing, the waveform will have much less intensity than nearby vowels or approximants, and the spectrogram will not have clear formants, especially above F3.

(C) Stop: Noted where there is a (larger, usually audible) burst, with or without voicing, and with or without aspiration. Usually there will be some period of closure (with or without voicing) before the burst, followed by aspiration or frication. This will sound like it has a much more aggressive onset that that of a regular fricative. There may be little or no aspiration after the fricative (as in a short-lag or voiced stop) or there may be aspiration and/or frication as in an aspirated stop. (But if the frication is very strong and/or long, it may be labeled as an affricate.)

(D) Approximant: A dental fricative realized as an approximant may sound vaguely /l/ -like. Formants may be clear or faint, but they have noticeably reduced amplitude relative to surrounding vowels. There should be no frication noticeable in the waveform or spectrogram. Flaps look like approximants that have slightly shorter duration and reduced amplitude, and especially unclear or absent formants, but there are many tokens that are ambiguous. Flaps are also labeled as approximants because there is no good way of distinguishing between them.

(E) Vowel: A dental fricative may be realized as a vowel, usually a /ə/ or /□/ type sound. The waveform is (nearly) indistinguishable from surrounding vowel sounds, and there are very clear formants, also (nearly) indistinguishable from surrounding vowels. The difference between the vowel and approximant is one of amplitude. Generally, approximants have reduced amplitude relative to vowels, and fainter formants. If there is

no decrease in amplitude, it is a vowel. If no trace of the dental fricative can be heard, then it is marked as a deletion. If there is a change in the vowel quality corresponding to the percept of a dental fricative, then the segment is labeled as a vowel.

(F) Nasal: It has antiformants, reduced amplitude, and is often next to another nasal. The nasality is clear in the spectrogram, but is also audibly obvious. If the combined segment is as short as a single segment, the dental fricative may just be deleted. If the nasal is long, the dental fricative is assimilated and realized as a nasal.

(G) Lateral: While approximant realizations may sound somewhat /l/-like, segments marked as laterals are obvious /l/s. These usually occur as complete assimilations to neighboring segments. If the original /l/ plus dental fricative as /l/ looks longer than a single segment, the dental is realized as a lateral. If the combined segment is as short as a single segment, the dental fricative may just be deleted.

(H) Affricate: A combination of a stop+fricative

(I) Mixed bag: If the sound has two distinct manner realizations, such as fricative+stop or fricative+approximant, both are noted.

PROSODY OF FOCUS AND CONTRASTIVE TOPIC IN K'ICHE'

Murat Yasavul*
Ohio State University

Abstract

This paper discusses the findings of an experimental study about the prosodic encoding of focus and contrastive topic in K'iche'. The central question being addressed is whether prosody plays a role in distinguishing string-identical sentences where the pre-predicate expression can be interpreted as being focused or contrastively topicalized depending on context. I present a production experiment designed to identify whether such sentences differ in their prosodic properties as has been impressionistically suggested in the literature (Larsen 1988; Aissen 1992; Can Pixabaj & England 2011). The overall strategy of the experiment was to obtain naturally occurring data from native speakers of

*I am indebted to the speakers of K'iche' who participated in this study and to Raul Castro, María Hernandez Us, Adelina Chom Canil and Juana Pérez Gómez for their judgments about the data I present. I also would like to thank Cynthia Clopper, Judith Tonhauser, Craige Roberts, Carl Pollard, Kathryn Campbell-Kibler, Mike Phelan, Laura Wagner, the participants of the Prosody-Semantics seminar in Fall 2010-Spring 2011 and the Prosody Working Group at Ohio State for their help with this study and for many helpful discussions about the material presented here. Of course, the usual disclaimers apply. I also thank Raul Castro, Heather Dean and Victoriano Canil for facilitating my fieldwork in Guatemala. The fieldwork for this project is funded by the Department of Linguistics and the College of Arts and Humanities at The Ohio State University.

Murat Yasavul. Prosody of focus and contrastive topic in K'iche'. *OSUWPL* Volume 60, Spring 2013, pp. 129-160.

K'iche' by having them repeat target sentences they heard in conversations. The phonological analysis showed that content words in K'iche' have a rising pitch movement, a finding which is in line with Nielsen (2005). The acoustic analyses of several variables yielded a significant effect of condition only in the range of the F0 rise associated with focused and contrastively topicalized expressions. However, the difference across conditions is only ∼6 Hz which may not be perceivable by listeners.

1 Introduction

In K'iche', a Mayan language of Guatemala, sentences like (1) may have two different interpretations given appropriate context (throughout, I use **boldface** for that part of the example which is relevant to the discussion at hand)[1]:

(1) **A** **Raul** x-∅-war-ik.
CLF Raul CMP-A3-sleep-SS
a. 'Raul slept.'
b. 'As for Raul, he slept.'

(1-a) is obtained when the pre-predicate expression *A Raul* 'Raul' is *focused*, i.e. when it is an answer to the Question Under Discussion (Roberts 1996), as in (1′)[2]:

(1′) Context: *Who slept?*
A **Raul**$_F$ x-∅-war-ik.
CLF Raul CMP-A3-sleep-SS
'Raul slept.'

(1-b), however, is obtained when the same expression is interpreted as a *contrastive topic*, the denotation of a topical constituent in a contrastive context (Roberts 2012), as in (1″):

[1]Unless otherwise stated, all the data in this paper are from original fieldwork in Santa María Tzejá, Ixcán, El Quiché, Guatemala and Columbus, Ohio, USA. In the orthography, all symbols have their standard phonetic value except the following: ' = glottal stop, C' = glottalized consonant, VV = long vowel, ch = [ʧ], tz = [ʦ], x = [ʃ], and j = [x] or [x̱]. The following abbreviations are used in the glosses of the examples: A1(p), A2(p), A3(p) = absolutive first, second, third person singular (plural) marker; E1(p), E2(p), E3(p) = ergative first, second, third person singular (plural) marker; 2s(p).f = second person singular (plural) formal; AFF = affectionate; AG = agent focus; AGT = agentive; AP = antipassive; ASP = aspect; CLF = classifier; COM = comitative; CMP = completive; COMP = complementizer; DAT = dative; DEM = demonstrative; DET = determiner; EMPH = emphatic; ENC = enclitic; FOC = focus particle; GEN = genitive; INCMP = incompletive; INSTR = instrumental; INTS = intensifier; IV = terminal suffix for morphologically intransitive verb; MOV = movement; NEG = negative particle; P>I = intransitive derived from positional; PART = particle; PERF = perfect; PL = plural; POS = positional; PREP = preposition; SS = status suffix; TOP = topic marker.

[2]The subscripts $_F$ and $_{CT}$ in K'iche' sentences indicate focused and contrastively topicalized expressions, respectively.

(1″) Context: A: *Raul and Roberto didn't work last night. Roberto went out.*
B: *And Raul, what did he do?*
A: **A** **Raul**$_{CT}$ x-∅-war-ik.
CLF Raul CMP-A3-sleep-SS
'As for Raul, he slept.'

A common property of focus and contrastive topic in K'iche', whose basic word order is predicate-initial, is that focused expressions may and contrastively topicalized expressions must be realized in the pre-predicate position. Additionally, such constituents can co-occur before the predicate as in (2), in which case the focused expression, here *al Maria* 'Maria', follows the contrastively topicalized expression, here *a Raul* 'Raul' (compare (2-a) and (2-b)). This fact provides language internal evidence that these discourse functions are distinguished by speakers.

(2) Context: *I know that Roberto saw Juana yesterday, but who did Raul see?*

a. A Raul$_{CT}$ al Maria$_F$ x-∅-r-il-o.
CLF Raul CLF Maria CMP-A3-E3-see-SS
'As for Raul, he saw Maria.'

b. #Al Maria$_F$ a Raul$_{CT}$ x-∅-r-il-o.
CLF Maria CLF Raul CMP-A3-E3-see-SS
(intended reading) 'As for Raul, he saw Maria.'

Alongside the change in basic word order, certain types of focus in K'iche' can be expressed by other morpho-syntactic means but these are neither obligatory nor do they apply across-the-board (more on this below). Consequently, this raises the question as to whether string-identical sentences like (1′) and (1″) differ in their prosodic properties because of the difference in their meaning. A broader question of interest is whether there always is a relation between pragmatics and prosody, in other words, whether meaning differences like the ones above are always reflected in the prosodic structure of otherwise identical sentences. Indeed, as regards K'iche', previous studies claimed that sentences like (1′) and (1″) differ in their prosodic properties. In particular, the literature has discussed whether the pre-predicate expressions in such sentences are set off from the rest of the sentence by a pause or not. Thomas Larsen (1991, p.c. cited in Aissen 1992) suggested that topics[3] in K'iche' are not followed by a pause. On the other hand, Can Pixabaj & England (2011) claimed that topics in K'iche', whether contrastive or not, are followed by a pause whereas foci are not.

In this paper, I discuss findings from a production experiment designed to identify whether the difference in meaning between focus and contrastive topic corresponds to a difference in prosody, which would distinguish string-identical sentences like (1′) and (1″) in K'iche'. The overall strategy of the experiment was to obtain naturally occurring data from native speakers of K'iche' by having them repeat target sentences they heard in conversa-

[3]There is no indication as to whether Larsen distinguished more than one kind of topic in K'iche'.

tions. The experiment was designed so that in each conversation, only one interpretation of the target sentence would be felicitous.

The rest of the paper is organized as follows. In section 2, I give the relevant background on K'iche' morpho-syntax which is necessary to understand the details of how focus and contrastive topic are expressed. In section 3, I summarize the literature on focus and contrastive topic in K'iche' in detail. I also elaborate on the differences between the current study and the previous work by making explicit my assumptions about focus and contrastive topic. After motivating the research question, I provide an overview of the previous work on the prosody of focus and contrastive topic in several languages including K'iche'. Section 4 presents the details of the production experiment, the analyses and the results before I conclude in section 5.

2 Background on K'iche' morpho-syntax

K'iche' is a Mayan language spoken by over a million people in the central and western highlands of Guatemala (Richards 2003). It has an ergative-absolutive agreement system (Larsen 1988) which is preserved throughout changes in aspect and clause type (Pye 2001). The basic word order is VS in intransitive clauses and VOA in transitive clauses (Larsen 1988; Pye & Poz 1988; England 1991), where S stands for the single argument of an intransitive, A for the more agent-like argument of a transitive, and O for the more patient-like argument of a transitive verb (Dixon 1994). In (3) and (4), I start with two examples that illustrate intransitive clauses.

(3) x-∅-war ri achi.
CMP-A3-sleep DET man
'The man slept.'

(4) x-at-war-ik.
CMP-A2-sleep-SS
'You slept.'

In K'iche', there is no case-marking on noun phrases, e.g. *ri achi* 'the man' in (3), to identify grammatical relations or semantic roles; these are read off of the verbal complexes via the ergative and absolutive cross-reference markers given in Table 1.

The absolutive markers are used to cross-reference, i.e. register the number and person features of, the S argument of an intransitive verb and the O argument of a transitive verb. In an intransitive verbal complex, e.g. *x-∅-war* 'CMP-A3-sleep' in (3), the sole argument *ri achi* 'the man' is cross-referenced by the phonologically null, third person singular absolutive marker -∅- 'A3' preceding the verb root *war* 'sleep'. The absolutive marker is preceded by the aspect marker *x-* 'CMP'. In (4), where the argument of the verb is not real-

Ergative	Preconsonantal	Prevocalic
E1	*-in-*	*-inw-/-w-*
E2	*-aa-/-a-*	*-aw-*
E3	*-uu-/-u-*	*-r-*
E1p	*-qa-*	*-q-*
E2p	*-ii-*	*-iw-*
E3p	*-ki-*	*-k-*

Absolutive	
A1	*-in-*
A2	*-at-*
A3	*-∅-*
A1p	*-uj-/-oj-*
A2p	*-ix-*
A3p	*-e'-/-eb'-/-ee-*

Table 1: Ergative and absolutive agreement markers

ized, the verbal complex also carries the status suffix *-(i)k* 'SS' following the verb root. This marker is claimed to mark phrase-finality, in particular, the end of an intonational phrase Henderson (2012) and it is used for intransitive verbs in the incompletive and completive aspects.

The other set of markers in Table 1, namely the ergative markers, are used to cross-reference the A argument of a transitive verb as exemplified in (5):

(5) x-at-u-to'-o.
CMP-A2-E3-help-SS
'S/he helped you.'

In a transitive verbal complex, e.g. *x-at-u-to'-o* 'CMP-A2-E3-help-SS' in (5), the absolutive marker *-at-* 'A2', which marks the O argument of the verb, precedes the ergative marker *-u-* 'E3', which marks the A argument. The ergative marker, in turn, precedes the verb root *to'* 'help'. Similar to intransitive verbs, transitive verbs may carry phrase-final suffixes when they occur at the end of intonational phrases (Henderson 2012). For example, in (5) the verb root is followed by the status suffix *-o*[4].

Since K'iche' does not use overt marking on noun phrases, it is the absolutive marker *-∅-* 'A3' that identifies, say, *ri achi* 'the man' as the O argument in (6) and it is the ergative marker *-u-* 'E3' that identifies *ri achi* in (7) as the A argument (Trechsel 1993):

(6) x-∅-a-to' ri achi.
CMP-A3-E2-help DET man
'You helped the man.'

(7) x-at-u-to' ri achi.
CMP-A2-E3-help DET man
'The man helped you.'

[4]The form of the status suffix for transitive verbs can be *-u*, *-o* or *-j* depending on the derivational status of the stem (Trechsel 1993). The status suffixes simultaneously register (in)transitivity, aspect and, in the case of transitive verbs, the derivational status of the stem (Pye 2001).

Although the basic word order in K'iche' is VS/VOA, in texts it is relatively uncommon to find the A, O or the S arguments in post-predicate positions realized as pronominal arguments. Larsen (1987:40) claims that independent pronouns, which rarely appear in argument positions, are used in some cases to indicate "contrastive emphasis" or change of subject[5]. These pronouns, given in Table 2 below, are identical to absolutive markers except for the third person singular and plural.

1sg	*in*
2sg	*at*
3sg	*are'*
1pl	*oj*
2pl	*ix*
3pl	*e a're'/a're'/ke*

Table 2: Pronouns in K'iche'

In addition to these pronouns, K'iche' marks formality/politeness for second person by two morphemes[6]: (i) *la* 'you' (singular) and (ii) *alaq* 'you' (plural) which occur post-verbally (Trechsel 1993). In (8), *la* '2s.f' marks the formal second person singular ergative argument and *oj* 'A3' marks the absolutive argument. In (9), *alaq* '2p.f' marks the formal second person plural argument:

(8) x-oj-to' **la**.
CMP-A3-help 2s.f
'You (sg. formal) helped us.'

(9) x-pee **alaq**.
CMP-come 2p.f
'You (pl. formal) went.'

The two word orders I discussed in this section characterize basic, non-emphatic sentences, i.e. sentences which do not involve topicalization or focus, and in which pronominal arguments are usually dropped. After this basic description of the relevant morpho-syntactic properties of K'iche', I now turn to the main topic of the paper, namely how the two discourse functions focus and contrastive topic are expressed.

[5]In the discussion of focus and topic below, we will see that these pronouns can occupy the pre-predicate positions when they are focused or topicalized.

[6]The formal pronouns will become relevant later on in the discussion of agent focus marking.

3 Focus and contrastive topic in K'iche'

3.1 Previous literature on focus in K'iche'

A general claim about Mayan languages, dating back to Norman (1977), is that they are generally predicate-initial, but that there are also two special positions preceding the predicate that constituents can occupy for pragmatic purposes. The discourse functions that these constituents have, which are called *focus* and *topic*, govern the changes in the basic word order in K'iche' (Larsen 1988; England 1991). Norman (1977) claimed that focus and topic are structurally different in that focus occupies the pre-predicate position whereas topic occurs sentence initially.

Focus constructions in Mayan have been traditionally analyzed as involving a movement operation whereby the focused constituent is realized in the pre-predicate position and linked to a gap in the post-focal portion of the sentence (Larsen 1988; Aissen 1992; Trechsel 1993)[7]. In her seminal work on topic and focus in Mayan, Aissen (1992) claimed that focused constituents occupy the [Spec, I′] position and bind a co-indexed trace lower in the tree. The constituents occupying the focus position are generally understood to be semantically "prominent" in some sense (Larsen 1988) as reflected in the cleft translation into English in (10), which is the standard practice in the Mayan literature (Aissen 1992; Larsen 1988; Trechsel 1993; Can Pixabaj & England 2011):

(10) **Aree ri achi** x-∅-q'ab'ar-ik.
FOC DET man CMP-A3-get.drunk-SS
'It was the man who got drunk.' (Larsen 1988:503)

Aissen (1992:43), in particular, claims that focus in Mayan has the two characteristics associated with the interpretation of clefts: an existence presupposition and a uniqueness assertion. The following example from Aissen (1992:49) is taken from the middle of a text in Tzotzil, where one individual, walking along, meets another working in a field who utters (11-a) and the narrative continues with (11-b). According to Aissen, in (11-b) there is a presupposition to the effect that there was something that the man was planting and that the focused expression *chobtik* 'corn' is the unique entity that satisfies this presupposition:

(11) a. 'I'm planting. I'm planting stones, I'm planting trees',
b. Pero **chobtik** tztz'un un.
but corn he.plants ENC
'But it was corn he was planting.' (Aissen 1992:49)

[7]In fact, Mayanists have traditionally subsumed pre-predicate focus constructions, content questions and relative clauses under the heading of focus because they characterized these constructions by the obligatory presence of a constituent preceding the predicate, the obligatory gap in the post-focal portion of the sentence, a dependency between them and the use of agent focus form (Larsen 1988; Trechsel 1993).

Focus constructions in Mayan are further characterized by a special verb form called the *agent focus* form, a much discussed phenomenon in the context of focus (see e.g. Mondloch 1981; Larsen 1988; Trechsel 1993; Aissen 2011 for K'iche' and Dayley 1981; Aissen 1999; Stiebels 2006 for other Mayan languages). Agent focus can only be used with transitive verbs when the ergative argument of the verb is focused as in (12):

(12) Aree le achi x-∅-kuna-**n** le ixoq.
FOC DET man CMP-A3-cure-AG DET woman
'It was the man who cured the woman.' (Trechsel 1993:42)

The verbal complex in (12), *x-∅-kuna-n* 'CMP-A3-cure-AG', is in the agent focus form which is expressed by (i) the absence of an ergative marker *-u-* 'E3' on the verb, and (ii) the presence of the agent focus marker[8] *-n* 'AG' attached to the verb. In (12), both the agent and the patient are third person singular and it is indeterminate whether it is the agent or the patient that the absolutive marker agrees with. Yet, when there is an agent focus marker, the interpretation is always that the pre-predicate argument, which denotes the agent of the action, is focused. Larsen (1988) points out that the agent focus form can never be used in simple transitive clauses.

In a recent study on K'iche' texts, Can Pixabaj & England (2011) argue that there are two types of focus in K'iche'. The first type is what they call *contrastive focus* which "usually requires an explicit contrast" and which, they claim, operates like clefts in English (p.23). In (13), for instance, Can Pixabaj & England (2011:22) say that the focused expression "explicitly contrasts 'my parents' with 'me', identified negatively in the previous clause ('it wasn't I who saw')":

(13) Pero **aree r-in-taat** k-e-tzjo-**n**-ik.
but FOC DET-E1-father INCMP-A3p-recount-AG-SS
'...but it was my parents who recounted (it).' (Can Pixabaj & England 2011:22)

According to Can Pixabaj & England, this kind of focus requires the use of the focus particle *aree* 'FOC' with definite nominals[9] as well as the agent focus form of the verb when ergative arguments are focused as in (13). They also claim that this type of focus is not followed by a pause (p.21)[10].

[8]This marker comes in two forms: *-(V)w* for root transitive verbs, and *-n* for derived transitive verbs (Trechsel 1993).

[9]Regarding the definite-indefinite distinction in K'iche', Can Pixabaj & England say "[w]e consider those that have no article or possessor, or have only the indefinite article *jun* to be "indefinite", while we consider those that are accompanied by one of the definite articles *wa, le, ri* (with or without the indefinite article), are possessed, are accompanied by demonstratives, or are proper names to be "definite"". They also claim that *xow* 'only' can precede definite nominals in contrastive focus contexts but they do not provide examples.

[10]The source of the examples in this study is based on five texts with more than 1,800 clauses. The commas in the texts after expressions in the pre-predicate position are taken to indicate pauses, and the lack thereof as evidence that there are no pauses.

The second type of focus that Can Pixabaj & England (2011:23) identify is used to "present new information", "mention a participant for the first time" or "reintroduce information". This type of focus is not used for "explicit contrast of old information", nor does it require the use of the particle *aree* 'FOC' or the use of the agent focus form. Yet, similar to the first type focus, focused expressions of this type are not followed by a pause (p.23). The following example, the first sentence of a recording, illustrates "mentioning a participant for the first time" where "the speaker is identifying the person who will speak, from a pool of all who are present" (p.23):

(14) Chanim, **le don Santiago** k-∅-u-tzijoj cha-q-e jas le
now DET don Santiago INCMP-A3-E3-tell PREP-E1p-DAT what DET
u'istoria r-ech we jun tinamit Santa Lu's.
E3-history E3-POS DET one town Santa Lucía
'Now don Santiago will recount the history of the town of Santa Lucia.'
(Can Pixabaj & England 2011:24)

The following is an example where the focus "reintroduces a participant", *ri achi* 'the man', which "was spoken of about 50 clauses ago, using *rajawal* 'master'" (p.24):

(15) es ke **ri achi**, ri r-ajaw w-u'lew rii', ∅-k'o jun u-tajkil
it.is that DET man DET E3-master DET-land DEM A3-exist one E3-errand
aw-uuk'
E2-COM
'...it is that the man, he who is the master of this land, has an errand with you...'
(Can Pixabaj & England 2011:24)

Can Pixabaj & England do not explicitly provide the contexts in which these sentences are uttered. So, for instance, in (15) we do not know what the immediately previous context is and, therefore, we do not know whether *ri achi* 'the man' is focused or topicalized. Similarly, in (14), we do not know why *don Santiago* is necessarily identified as the person who will speak. This sentence may very well be an all focus sentence answering an implicit question like *What is going to happen now?* In fact, later on in the paper, Can Pixabaj & England (2011:26) note that this type of focus has the same *function* as a (non-contrastive) topic and the only difference between them is that the latter is followed by a comma in their textual data. All in all, given the lack of explicit contexts and definitions, it is hard to assess Can Pixabaj & England claims.

To summarize, we have seen that focus in K'iche', just as in other Mayan languages, can occur before the predicate. According to Aissen (1992), Larsen (1988) and Trechsel (1993) focus sentences are interpreted like clefts in English. According to Can Pixabaj & England (2011), however, K'iche' focus divides into two and only those sentences where the focused expression is preceded by *aree* 'FOC' are interpreted like clefts. Can Pixabaj

& England further claim that foci in K'iche', regardless of their type, are not followed by a pause. In the following section, I will elaborate on the assumptions I am making about focus and in doing so show how they can be applied to K'iche'. These assumptions are necessary to elaborate on my research question.

3.2 Background assumptions about focus

At an intuitive level, focus involves a way to mark "highlighted" or "emphasized" information in discourse. This seems to have been the general approach in the Mayan community in terms of its characterization of what is meant by the term focus. Despite ample discussion of this phenomenon in the literature along similar lines, the present study makes different assumptions about focus and how focus is expressed in K'iche'. Part of the reason for this departure from the common assumptions is empirical in that the generalizations made in the literature do not hold up against the data that I collected, which I illustrate below. Yet, the main motivation for a different characterization of focus in K'iche' is to situate it in the broader semantic-pragmatic literature and to have a principled characterization of focus that makes predictions. It might turn out that these assumptions need to be revised but the advantage of the framework that I will summarize below is that it gives us working definitions that we can test. It is not always clear what is meant by "new information", "emphasis" or "reintroducing a participant" etc. and without explicit definitions of these discourse functions, it is hard to come up with adequate analysis of the pragmatic phenomena that are under discussion.

A more principled characterization of focus is to consider it as answering an explicit or implicit question (Jackendoff 1972; Roberts 1996), which Kadmon (2001) claims to be the most basic and crucial intuition about focus. From this point of view, in a constituent question-answer pair, the phrase corresponding to the question-word is focused. So, for instance, in English the sentence *Michael ate tortillas*, with prosodic prominence on *Michael*, in particular a H* accent followed by a L-L% boundary tone[11] (Jackendoff 1972; Büring 2003) can constitute a felicitous answer to *Who ate tortillas?* (16) but not to, say, *What did Michael eat?* (17):

(16) Context: *Who ate tortillas?*
MICHAEL ate tortillas.
H* L-L%

[11]The letters L and H are used in the Autosegmental-Metrical (AM) Framework (Pierrehumbert 1980), which is a framework for intonational analysis. In the AM theory, the prosodic grouping and prominence relations are represented by distinctive pitch events, transcribed by a sequence of Low (L) and High (H) tones. or combinations thereof. The tones are marked with diacritics indicating their intonational function. There are *pitch accents* that mark prominence and *boundary tones* that mark the edges of prosodic boundaries. A star (*) on a pitch-accent indicates that it is associated with a stressed syllable. The % sign on a tone indicates a prosodic boundary.

(17) Context: *What did Michael eat?*
#MICHAEL ate tortillas.
H* L-L%

In (16), where the Question Under Discussion (QUD) is *Who ate tortillas?*, *Michael* corresponds to *who* whereas the rest of the sentence, *ate tortillas*, is congruent to the QUD in the sense that abstracting on the *wh*-word in the question yields the property $\lambda_x.x$ ate tortillas which is also the denotation of the rest of the sentence. Sometimes this partitioning with respect to a QUD is termed as the *Theme/Rheme* distinction where *Rheme* denotes the focus and *Theme* denotes the part of the sentence congruent to the QUD (Roberts 2012). A QUD is a semantic question, i.e. a set of propositions, that corresponds to the current discourse topic (Roberts 1996:93). It may be the denotation of an actual question that is asked as in (16) above or may be implicit in the discourse (Roberts 1998). As the examples above illustrate, focus presupposes that there is such a QUD, a presupposition which, together with contextual clues, enables the addressee to reconstruct, or *retrieve*, the QUD (Roberts 1996).

A related and widely-held view about focus is that it evokes alternatives in discourse (Rooth 1992). According to Rooth's analysis of focus interpretation, prosodic prominence on *Michael* in (16) evokes alternatives such as *Robert*, *Jane*, *Peter*, etc. with which one constructs a set of propositions of the form, x ate tortillas, for the original sentence where x ranges over possible alternatives drawn from a contextually restricted set E. This set of alternatives that focus evokes helps determine an additional semantic value for an utterance, which Rooth calls *the focus semantic value*. In other words, the focus semantic value of a focused expression α, denoted by $[\![\alpha]\!]^{\mathrm{f}}$, is obtained by making a substitution in the position corresponding to the focused expression in the sentence. To illustrate, the focus semantic value of (16) is given in (18). The ordinary semantic value can be drawn from the focus semantic value as the former is always an element of the latter (Rooth 1992:76). Crucially, the focus semantic value of (16) is the same set we obtain by abstracting on the *wh*-word in the question in (16), hence the question-answer congruence (Roberts 1996).

(18) $[\![[\text{MICHAEL ate tortillas}]]\!]^{\mathrm{f}}=\{\mathbf{ate}(x, \mathbf{tortillas})|x \in E\}$

So far, I have illustrated the question-answer congruence with English examples where focus is marked prosodically. Yet, Roberts (1996) points out that the prosodic realization of focus is not universally assumed by those working on the semantics of focus. This means that focus may involve non-prosodic means and, in fact, many languages use cleft-like structures, marked word order or special morphemes to indicate focus in addition to intonational marking (Büring 2011). Therefore, the common core of focus is the observation that it evokes alternatives and that it is intuitively linked to question-answer congruence irrespective of the actual means of realizing focus (Roberts 1996; Rooth 1996).

In the present study, I follow the line of thinking summarized above and characterize

focus in K'iche' as follows: (i) a focused expression can occur before the predicate and (ii) its meaning will yield an answer to the QUD when the meaning that is congruent to the QUD applies to it. In other words, an answer to the QUD, say in (16), is obtained by applying the property $\lambda_x.\ x$ ate tortillas to the meaning of the focused expression. The question-answer congruence, the defining characteristic of focus, can be shown in K'iche' as follows. Consider the examples in (19) and (20): (19-a)/(20-b) is a felicitous answer in (19) but not in (20), and (20-a)/(19-b) is a felicitous answer in (20) but not in (19):

(19) Context: *Who helped you?*

a. **A** **Raul**$_F$ x-in-u-to'-o.
CLF Raul CMP-A1-E3-help-SS
'Raul helped me.'

b. #**In**$_F$ x-in-u-to'-o.
I CMP-A1-E3-help-SS
'He helped me.'

(20) Context: *Who did Raul help?*

a. **In**$_F$ x-in-u-to'-o.
I CMP-A1-E3-help-SS
'He helped me.'

b. #**A** **Raul**$_F$ x-in-u-to'-o.
CLF Raul CMP-A1-E3-help-SS
'Raul helped me.'

I will end this section by discussing two properties of the focus stimuli that I used in the experiment. Recall that the research question of the present study builds on the observation that focus and contrastive topic sentences can be string-identical. In order for this to hold, the focus sentences should not carry any special focus marking except for the change in word order because contrastive topics, as we will see below, occur before the predicate with no additional morpho-syntactic marking. Consequently, none of the focus stimuli had the focus particle *aree* 'FOC' in them and, furthermore, when ergative arguments were focused, the agent focus marker wasn't used.

Although it is widely discussed as a concomitant of focusing ergative arguments, the agent focus marker was not obligatory for my informants and they were not making use of this form very often in elicitation sessions. Larsen (1988:505) also reports that using agent focus is optional even when its use is permissible. In any case, there are restrictions regarding the use of agent focus. For instance, at least one of the arguments of the verb has to be third person or second person formal for the use of agent focus to be felicitous:

(21) *__In__ x-at-ch'ay-__ow__-ik.
I CMP-A2-hit-AG-SS
(intended reading) 'I hit you.'

In order to focus the agent NP in (21) one can: (i) use the active voice as in (22), or (ii) demote the patient NP and use the oblique phrase *aw-e* 'E2-GEN' as in (23):

(22) **In** x-at-in-ch'ay-o.
I CMP-A2-E1-hit-SS
'I hit you.'

(23) **In** x-in-ch'ay-**ow** **aw-e**.
I CMP-A2-hit-AG E2-GEN
'I hit you.'

In summary, the agent focus form cannot always be used. Even when it is applicable, there are either restrictions on its use or it alternates with the active form of the verb. Moreover, there is no counterpart of agent focus for intransitive verbs or for cases where the O-argument of a transitive verb is focused. When these arguments are focused, the active form of the verb is used (Larsen 1988; Trechsel 1993). This shows that in general foci can be marked by only a change in word order just as topics. In the next section, I turn to the discussion of topics in K'iche'.

3.3 Previous literature on topics in K'iche'

The second discourse function that can be expressed by a pre-predicate expression in K'iche' (and in Mayan languages in general) is called *topic*. The topic of a sentence is defined to be the constituent that indicates what the sentence is about (Aissen 1992; Roberts 2012) and in this sense a topicalized expression is an entity to which our attention is drawn (Aissen 1992; Roberts 2012). Below is an example in K'iche' where the topicalized expression *Ri ulew* 'the earth' precedes the predicate:

(24) Context: *Tell me something about the earth.*
Ri **ulew** k-∅-b'in chi-r-ij ri q'iij.
DET earth INCMP-∅-walk PREP-E3-around DET sun
'The earth revolves around the sun.'

Aissen (1992) distinguishes between two kinds of topics in Mayan languages: (i) *external topics* and (ii) *internal topics* and argues that these topics behave differently both structurally and pragmatically. The following example from Tzotzil illustrates external topics. Aissen points out that the first line in (25) introduces two discourse participants, the second line turns attention to one of them, namely *a ti vinik-e* 'the husband' and asserts something about him, and the third does the same for the other participant, here *a ti antz-e* 'the wife'. Both of the topicalized expressions are preceded by the topic marker *a* 'TOP'. They are also usually accompanied by a definite determiner *ti* 'DET' and an enclitic *e* 'ENC'.

(25) a. There was a man and a woman, newlyweds.
b. **a ti vinik-e** ta-xlok' ech'el, ta-tbat ta-xxanav.
TOP DET man-ENC exists away goes travels
'The husband leaves, he goes, he travels.'
c. **a ti antz-e** jun-yo'on ta-xkom
TOP DET woman-ENC happily stays
'The wife stays at home happily...' (Aissen 1992:49)

Aissen refers to external topics as new or shifted topics: once a participant is topicalized in this way, it is not referred to again by an overt nominal unless the topic shifts to another participant (p.51). Structurally, external topics occupy a position outside the clause, as a sister of the CP, and are base-generated. There is no requirement that they bind a coreferential pronoun lower in the clause[12]. Their structure, therefore, resembles that of left-dislocation (p.48) where the topic is prefixed to a fully well-formed root CP so long as the CP is about the topic. Aissen also makes a claim about the prosody of such topics and says they are followed by a pause, which, in her theory, follows from the syntactic structure (p.76).

The second kind of topic Aissen identifies, namely internal topics, involves discourse participants which are already identified as topic and can occur in the pre-predicate position. The following piece of discourse in Tz'utujil provides an example of such topics. The texts starts '[a] long time ago there was a man whose daughter was in a dance' and Aissen claims that (26-a) introduces *rme'al* 'his.daughter' as a new topic, marked by the particle *ka'(ar)* 'PART', and this same topic is referred to again by an overt nominal in the following sentence in (26-b):

(26) a. **Ja k'a rme'al** x-u-koj pa xajoj xin Tukun.
the PART his.daughter ASP-E3-enter in dance of Tecun
'He entered his daughter in the dance of Tecun.'
b. y **ja rme'al** x-ok-i Malincha.
and the his.daughter ASP-play-IV Malincha
'and the daughter played the part of the Malincha.' (Aissen 1992:74-75)

Such NPs can occur in the topic position although their referent has already been established. Structurally, these topics occupy the [Spec, C′] position, and like foci, bind a co-indexed trace lower in the clause. Furthermore, these topics are not separated from the following clause by a pause. As regards K'iche', Thomas Larsen (p.c. 1991, as cited in Aissen 1992) suggested that topics in K'iche' have the function of external topics in terms of their meaning but are associated with the syntax of internal topics, i.e. there is no pause after them.

[12]In Jakaltek, such topics may bind overt pronouns in the CP and yet in Tzotzil we don't find these pronouns as the language is pro-drop (Aissen 1992:69).

Building on Aissen's work, Can Pixabaj & England (2011) argue that there are two types of topics in K'iche'. Their characterization of topics is structural in that they are interested in "defining structurally the preverbal positions that can be filled by noun phrases". According to their characterization, the first type of topic occurs in "the first position" (sentence-initial; MY) preceding the verb and has no "special" marker such as *aree* 'FOC' or "special" verb form (agent focus form; MY) when it is the subject of a transitive verb (p.19). An example is (27) where Can Pixabaj & England claim that the hunter "was introduced in the previous clause and is here established as the local topic and continues as such for three more clauses, with only anaphoric reference" (p.20):

(27) **Ri k'aq-an-eel**, iii b'yeen ∅-u-b'an-om k'ax ch-k-e
DET hunt-AP-AGT eh INTS A3-E3-do-PERF bad PREP-A3p-DAT
s-taq-a'waj-iib'.
AFF-PL-animal-PL
'The hunter had done much damage to the animals.'
(Can Pixabaj & England 2011:20)

As with the examples of the different kinds of focus above, Can Pixabaj & England do not provide the context in which this sentence is uttered. Therefore, we do not know how the expression *ri k'aqaneel* 'the hunter' was introduced and whether it occurred in the pre- or the post-predicate position nor do we know whether it was focused or topicalized in the previous clause.

The second type of topic that Can Pixabaj & England (2011) identify is called *contrastive topic* which combines the functions of topic and focus "in the context of changing the topic and at the same time contrasting it with the previous topic" (p.24). According to Can Pixabaj & England, such topics can be preceded by the phrase *aree k'u*[13]. Unlike "contrastive focus", however, there is no "special" verb form (agent focus form; MY) that can be used with this construction. Furthermore, the nominal which is contrastively topicalized is followed by a pause. An example that Can Pixabaj & England (2011) provide is (28) where they say that in clauses before this example "the topic was the hunter, now it is the master of the mountain where he went to hunt" (p.24):

(28) Tonse are k'u **ri r-ajaw-al** u-winaq-il ri' ri jyub',
well EMPH PART DET E3-master-ABST E3-person-ABST DEM DET ill
jawi r-qas -k-∅-e'-k'aqa-n-a wi, x-∅-tak'-i'
where DET-always INCMP-A3-MOV-hunt-AG-SS EMPH CMP-A3-standing-P>I
r-oyowaal.
E3-anger
'Well, on the other hand the master of the hill, where he always went to hunt, got mad.' (Can Pixabaj & England 2011:25)

[13]A similar claim made by López Ixcoy (1997) is that the particle *aree* 'FOC' itself precedes contrastively topicalized expressions.

Given that Can Pixabaj & England take topics to occur only before the predicate, it is safe to assume that *the hunter* occurs before the predicate in the clause preceding (28). Yet, we do not know why *the hunter* was the topic rather than the focus because Can Pixabaj & England do not provide the context in which the sentence is uttered. This lack of contextual evidence makes it hard to determine the discourse status of the pre-predicate expressions in the examples they present.

To summarize, we have seen that in Mayan topicalized expressions occur before the predicate and generally two kinds of topic are distinguished. Aissen's external topics and Can Pixabaj & England's contrastive and non-contrastive topics are all separated from the post-topical portion of the sentence by a pause. In the next section, I summarize the assumptions I am making about topics in K'iche' and point out the differences between the previous literature and the present study.

3.4 Background assumptions about topics

In this section, I will present the assumptions I am making about topics in K'iche' and the kinds of topical constituents that are realized before the predicate. The first kind of topic, which can be realized in the pre-predicate position, indicates what the sentence is about (29):

(29) Context: *What happened to Raul?*
A **Raul** x-∅-tzaq-ik.
CLF Raul CMP-A3-fall-SS
'Raul fell.'

The other kind of topical constituent, which, as far as my data suggest, is always realized in the pre-predicate position, behaves as a contrastive topic, i.e. the denotation of a topical constituent in a contrastive context (Roberts 2012). A contrastive topic, alongside being a topic, also implies that there is another question about a different topic. Put differently, it implies that there are other entities having the same type as the contrastively topicalized expression and that we are going through a list, so to speak, and answering the QUD with respect to the entity at hand. Consider the example below where, when the topic changes from *Raul* to *Roberto*, the new sentence answers the question with respect to *Roberto*:

(30) Context: A: *Raul and Roberto are farmers. Last year, Raul sowed corn.*
B: *And Roberto, what did he sow?*
A: **A** **Roberto**$_{CT}$ x-∅-u-tik kinaq'.
CLF Roberto CMP-A3-E3-sow beans
'As for Roberto, he sowed beans.'

Contrary to what Can Pixabaj & England (2011) and López Ixcoy (1997) claim, my consultants did not accept neither the marker *aree k'u* nor the marker *aree* with contrastively topicalized expressions. However, these markers were acceptable for them when the pre-predicate expression was focused. Consequently, contrastively topicalized expressions in my data do not carry the markers *aree* or *aree k'u* and, therefore, can be string-identical to focus sentences without *aree* and without agent focus marking.

3.5 Previous studies on the prosody of focus and contrastive topic

So far, I have shown how focus and contrastive topic are expressed in K'iche' and how a sentence with a pre-predicate focus can be string-identical to a sentence with a contrastive topic. I have also noted that the literature on K'iche' has discussed whether the focused or contrastively topicalized expression is set off from the rest of the sentence by a pause. There have been two opposite claims with respect to this issue: (i) Thomas Larsen (1991, p.c. cited in Aissen 1992) claims that K'iche' topics function as external topics in the sense of Aissen (1992) but are not followed by a pause, and (ii) Can Pixabaj & England (2011) claim that topics in K'iche' are followed by a pause regardless of their type whereas foci are not followed by a pause regardless of their type.

Whether there is a pause or not following pre-predicate expressions is one potential prosodic cue that is of interest to the present study. On the other hand, work on other languages has suggested that there are other prosodic cues associated with focus and contrastive topic which need to be taken into account in a thorough study of the prosody of focus and contrastive topic. In the following sections, I briefly summarize the cross-linguistic findings about the prosodic encoding of focus and contrastive topic. I first start with a summary of the work in languages other than K'iche' and then turn to the details about K'iche'.

3.5.1 Previous studies on other languages

It has been shown that prosodic prominence on a focused expression can be indicated by various phonological and phonetic means. In English, for example, it has been claimed that focus is primarily marked by a pitch accent, in particular by a H* pitch accent followed by a L-L% boundary tone (Jackendoff 1972; Büring 2003). In fact, it is argued that this intonational contour distinguishes focus from contrastive topic in English as the latter is marked by a L+H* pitch accent followed by a L-H% boundary tone (*ibid.*)[14]. In general, accenting has been taken as the primary source of prosodic prominence marking, at least for English (Rooth 1992; Kadmon 2001; Féry & Samek-Ladovici 2006).

For languages other than English, research has shown that prosodic prominence on a focused expression may be realized through a variety of phonetic and phonological

[14]The accents marking focus and contrastive topic have also been called A and B accent, respectively (Jackendoff 1972), and fall and fall-rise accent, respectively (Büring 2003).

means. For example, in some languages, e.g. Italian (Grice *et al.* 2005) and Spanish (Face 2002), different pitch accents are used to indicate focused expressions. Yet, in some other languages, e.g. Korean (Jun 2005) and Japanese (Venditti *et al.* 2008), prosodic prominence is realized through phrasing, namely by placing a prosodic phrase boundary before or after the focused expression to indicate prominence. In these languages dephrasing can be used to mark expressions as less prominent, which is similar to the use of deaccenting in English. These various phonological properties show that, cross-linguistically, different means are available to indicate prosodic prominence, e.g. accenting, phrasing.

Alongside these phonological means, many languages indicate prosodic prominence through phonetic means. For example, focused expressions in English are typically longer in duration (Cooper *et al.* 1985) and have an expanded pitch range compared to non-focused expressions (Eady *et al.* 1986). Similarly, in Mandarin, focused expressions have an increased pitch range and the pitch range of the post-focal expressions is compressed (Wang & Yu 2011). Another phonetic cue to prosodic prominence involves the alignment of the pitch accent peak. In Spanish, the alignment is earlier (Face 2001) whereas in German it is later on a focused expression (Braun 2006) compared to a non-focused expression.

As regards the prosody of contrastive topic, research has shown that such expressions have a particular prosodic structure, too. I have already noted above that contrastive topics in English are marked by a L+H* pitch accent followed by a L-H% boundary tone. In German, contrastively topicalized expressions carry a late-rising pitch accent and are prosodically separated from the main clause by a prosodic boundary (Féry 2006). In Mandarin, topics raise the initial pitch range but there is no prosodic correlate of contrastiveness of topics (Wang & Yu 2011).

In sum, prosodic effects of focus and contrastive topic can be indicated through both categorical phonological means and continuous phonetic means. An adequate study of the prosodic reflexes of discourse functions like focus and contrastive topic should take such means into account in the analysis.

3.5.2 Previous studies on K'iche'

Although the phonology of K'iche' is well-described (Mondloch 1978; López Ixcoy 1997; Larsen 1988), there are not many studies dedicated to its prosodic structure. Nevertheless, there have been some claims about the prosody of focus that I will present in this section.

A study devoted to a preliminary prosodic description of K'iche' is Nielsen (2005). Nielsen's work is different from all of the other work on K'iche' that makes claims about prosodic structure in that it involves intonational analyses of utterances from a native speaker rather than impressionistic claims or text analysis. In her study, Nielsen found that K'iche' has stress driven pitch accent and L+H* is the default pitch accent on content words. This finding is in line with the previous literature which claimed that K'iche' has word-final stress (Larsen 1988). Nielsen also described K'iche' as an accentual phrase

language where prosodic domains which may be slightly larger than a word, namely *accentual phrases*, are marked by a tone. According to Nielsen, the default L+H* accent on the prominent syllable of a content word also marks the boundary of an accentual phrase. Alongside these findings about the general prosodic structure of K'iche', Nielsen found upstepped pitch accents associated with focused expressions where the L tone of the L+H* associated with the focused expression starts higher than the previous L.

The other claims about the prosody of focus are related to the interaction between focus and negation. It has been traditionally claimed that negation in K'iche' is indicated by the negative particle *man*[15] before the predicate and the so-called irrealis particle *ta(j)*[16] after the predicate, with the form of *ta(j)* changing depending on where it occurs (Larsen 1988; López Ixcoy 1997; Can Pixabaj 2010; Henderson 2012). Henderson (2012) claims that the distribution of *taj* is the same as the status suffixes *-(i)k* and *-o*, i.e. it occurs at the end of intonational phrases[17]. Examples (31-b) and (32-b) below, which are the negated versions of (31-a) and (32-a), respectively, illustrate this variable pattern. In (31-b), *ta(j)* occurs at the end of an intonational phrase and is realized as *taj* whereas in (32-b) it assumes its non-phrase-final form and is realized as *ta* (Larsen 1988; Henderson 2012):

(31) a. X-∅-war-ik.
CMP-A3-sleep-SS
'S/he slept.'

b. **Man** x-∅-war **taj**.
NEG CMP-A3-sleep NEG
'S/he didn't sleep.'

(32) a. X-∅-inw-il ri achi.
CMP-A3-E1-see DET man
'I saw the man.'

b. **Man** x-∅-inw-il **ta** ri achi.
NEG CMP-A3-E1-see NEG DET man
'I didn't see the man.'

Larsen (1987:51) claims that when focus constructions are negated, the negation

[15]It has been reported that the negative particle *man* exhibits dialectal variation. In some dialects it is *man*, in some dialects it is *ma* and in yet others it is *na* (Larsen 1988; Henderson 2012).

[16]This particle has been traditionally glossed as an irrealis particle in K'iche' and it does have an irrealis meaning when it is used in counterfactual constructions Larsen (1988). However, it can be used without *man* in a negated sentence because, as Larsen points out, in many dialects of modern K'iche', the negative particle *man* is optional. In the speech of all but one of the consultants that I worked with, *man* is almost always omitted and only the so-called irrealis particle *ta(j)* is used. I, therefore, follow (Pye 2001) and treat *ta(j)* as a negation particle and gloss it as NEG in negated sentences.

[17]In the speech of my consultants, the non-phrase-final form *ta* is always realized in a reduced form as [t] cliticized to the preceding word. See Romero (2012) for a similar observation about the phonological realization of this particle.

particles are placed around the focused expression and, in particular, in the negated (33-b), the negation particle assumes its phrase-final form *taj* (translations are Larsen's):

(33) a. Are' x-∅-ch'ay-ow ri achi.
he CMP-A3-hit-AG DET man
'He was the one who hit the man.'
b. **Man** are' **taj/*ta** x-∅-ch'ay-ow ri achi.
NEG he NEG CMP-A3-hit-AG DET man
'He was not the one who hit the man.' Larsen (1987:51)

A conclusion that Larsen draws by comparing (33) to (34), where *taj* occurs before a clause boundary, is that since the focused expression in (33) is followed by the phrase-final form *taj*, there is a clause boundary immediately before the verbal complex showing that the focused constituent is separated from the post-focal material[18].

(34) Le achi **ma** x-∅-uu-chooma-j **taj** chi x-in-aa-ch'ay-o.
DET man NEG CMP-A3-E3-think-SS NEG COMP CMP-A1-E2-hit-SS
'The man didn't think that you hit me.' Larsen (1987:50)

However, Henderson (2012) claims that focused expressions form a phonological phrase in K'iche' and not an intonational phrase (pp.19-20). Therefore, the focused constituent cannot be followed by the phrase final form *taj* but rather by *ta*, the non-phrase final form of the negation particle:

(35) **Man** are' **ta(*j)** x-∅-r-il-o.
NEG s/he NEG CMP-A3-E3-see-SS
'S/HE didn't see him/her. Henderson (2012:19)

Before going into the details of the experiment, I will mention some relevant points about the claims we have seen so far for the study at hand. For instance, if focused expressions in the data are followed by intonational phrase boundaries, then that will provide counterevidence for Henderson's (2012) claim that focused constituents do not form intonational phrases. If, on the other hand, the pre-predicate expressions are not followed by pauses then that would provide counter-evidence to the claim put forth in Can Pixabaj & England (2011) that topics in K'iche' are followed by a pause. Since the focused expressions are not preceded by any other expression in the experimental stimuli, it is not possible to test whether Nielsen's (2005) claim about up-stepped pitch accents holds true for my data. Yet, it is possible to see if her description of the prosodic structure of K'iche' is reflected in the data I collected. These and the claims about the prosodic reflexes of focus and contrastive

[18]Larsen's claim is not necessarily a prosodic one as he conceives of the boundary as a syntactic clause boundary (p.51).

topic in other languages will be taken into account in the prosodic analyses of the data.

4 The experiment

The discussions in the previous sections provide support for the claim that K'iche' allows string-identical sentences to have different interpretations as in (36) repeated here from section 1:

(36) **A Raul** x-∅-war-ik.
CLF Raul CMP-A3-sleep-SS
a. 'RAUL slept.'
b. 'As for Raul, he slept.'

This raises the question as to whether such sentences differ in their prosodic properties. In order to answer this question, I designed and carried out a production experiment with native speakers of K'iche' which aimed to obtain naturally occurring data. In a nutshell, the experiment involved participants listening to conversations accompanied by visual stimuli. The last sentence of each conversation was a target sentence where the pre-predicate expression was interpreted either as a focus or a contrastive topic depending on context. The task for each participant was to utter the target sentence as an answer to a question that is part of the conversation s/he heard. The following sections lay out the details of this experiment.

4.1 The participants

The experiment was carried out with 6 (4F, 2M) native speakers of the Joyabaj dialect of K'iche' in Santa María Tzejá, Ixcán, Guatemala in the summer of 2011. All of the participants were bilingual in K'iche' and Spanish and non-literate in K'iche'. They did not report any hearing, speech or visual impairments. The speakers were paid for their participation in the study.

4.2 Methods

4.2.1 The stimuli

The stimuli of the experiment consisted of 32 context-target utterance sequences (16 contrastive topic, 16 focus) and 19 fillers consisting of similarly constructed discourses. An example discourse for a focus sentence, which consists of a question-answer pair, is given in (37):

(37) A: Chin x-∅-u-b'an le wa?
who CMP-A3-E3-make DET tortillas
'Who made the tortillas?'
B: **Al Maria**$_F$ x-∅-u-b'an le wa.
CLF Maria CMP-A3-E3-make DET tortillas
'Maria made the tortillas.'

The corresponding discourse where the pre-predicate expression is contrastively topicalized is given in (38). Here, speaker B introduces two discourse participants, *Maria* and *Manuela*, and says something about *Manuela*. Speaker A then asks about *Maria*, and B's answer to that question is the target sentence which is string-identical to the one in (37). Note that in the case of contrastive topic, the stimulus is not just a question-answer sequence but rather has an additional sentence which introduces two discourse participants before the question is asked.

(38) B: Al Maria r-ichbil al Manuela x-∅-ki-b'an rikil.
CLF Maria E3-and CLF Manuela CMP-A3-E3p-make dinner
'Maria and Manuela made dinner.'
Al Manuela x-∅-u-b'an le kinaq'.
CLF Manuela CMP-A3-E3-make the beans.
'Manuela made the beans.'
A: Al Maria, su x-∅-u-b'an-o?
CLF Maria what CMP-A3-E3-make-SS
'Maria, what did she do?'
B: **Al Maria**$_{CT}$ x-∅-u-b'an le wa.
CLF Maria CMP-A3-E3-make DET tortillas
'As for Maria, she made the tortillas.'

As in the examples above, each discourse was constructed in a way to make only one interpretation of the target sentence possible. The target utterance was always the last sentence of a given discourse and was always an answer to a question eliciting a focused or contrastively topicalized expression. All of the pre-predicate expressions in the target sentences were proper names with penultimate stress. This enabled me to make the pre-predicate argument long enough to be able to clearly observe any associated intonational event. Furthermore, the target utterances had the same number of syllables in all of the stimuli. In this way, I ensured that any differences observed in the prosody were due to a difference in information structure.

Almost all of the people living in Santa María Tzejá where the experiment was carried out are non-literate in K'iche'. Therefore, it was not possible to use written material in the design of the experiment. Rather, the stimuli were recorded as conversations that the participants could listen to. In order not to bias the participants with native speaker

prosody, all of the target sentences were recorded by two non-native speakers of K'iche'. The questions, on the other hand, were recorded by a native speaker of K'iche'. Hence, all of the conversations were between a native and a non-native speaker of the language. To reduce the memory load, each conversation was accompanied by visual stimuli, e.g. pictures of women preparing beans and tortillas for the examples in (37) and (38). For each target sentence, the auditory and visual stimuli were the same across conditions. Figures 1 and 2 show the setup used for the contexts given above.

Maria

Figure 1: An example visual stimulus for focus

Manuela

Maria

Figure 2: An example visual stimulus for contrastive topic

4.2.2 The procedure

The participants were told that they were going to listen to conversations that consisted of question-answer pairs between a native speaker and two non-native speakers of K'iche'. They were also told that the non-native speakers were interested in hearing how native speakers would say the answers in the conversations. In Figures 1 and 2, clicking on the loudspeaker on the left played the conversation as a whole and clicking on the loudspeaker on the right played the same conversation without the target sentence. The task for each participant was first to listen to each conversation as a whole 1-2 times. Then the participant would listen to the same conversation one more time where the target sentence was removed and repeat the last sentence of the initial conversation as an answer to the question asked in the second conversation.

The participants were seated at a table in front of a laptop. Each participant wore head-mounted Sennheiser HMD280 headphones with microphone. The recordings were made with an Edirol R-09 recorder. 26 out of 192 utterances were excluded due to disfluency and the remaining 166 were included in the prosodic analysis.

4.3 Results

The research question I started out with was whether string-identical sentences with different meanings, namely focus and contrastive topic, also differ in their prosodic properties. The previous literature on contrastive topic and focus in K'iche' discussed whether constituents bearing these discourse functions are set off from the rest of the sentence by a pause. In order to see if this claim holds for the data I collected, each utterance was divided into two parts: (i) the pre-predicate part and (ii) the post-focal or post-topical part. I start with a discussion of the prosody of the pre-predicate expressions.

In all of the target utterances, the pre-predicate expression contained a rising pitch movement associated with the stressed syllable of the proper name. This finding is in line with the previous literature, in particular with Nielsen (2005) who claimed that K'iche' has stress-driven pitch accent where L+H* is the default pitch accent on content words.

In 50 (out of 166, ≈30%) utterances, the pre-predicate expression was followed by a pause. In the analysis, any physical pause between the pre-predicate expression and the rest of the sentences was taken into consideration. Such a pause after the pre-predicate expression was a proper pause and did not involve a stop closure because the following expression was always a verbal complex which began with a [ʃ] (*x* in the K'iche' orthography which stands for the completive aspect marker). The 50 pauses were distributed among the two conditions as follows: 29 focus, 21 contrastive topic. The mean duration of the pauses was 0.126s for focus and 0.115s for contrastive topic. A linear mixed effects model with speaker and item as random variables and condition as an independent variable did not yield any significant effect of condition. This finding shows that there is no clear indi-

cation that contrastive topics are distinguished from foci by a pause following them which goes against the claim made by Can Pixabaj & England (2011). As regards the prosodic boundaries following pre-predicate expressions, 77 of the focused expressions (out of 83, 92.7%) and 76 of the contrastively topicalized expressions (out of 83, 91.5%) were marked by a H% boundary tone. This finding shows that the boundary tone following the pre-predicate expressions is not affected by condition. It also goes against the claim that focus constituents do not form their own intonational phrases (Henderson 2012).

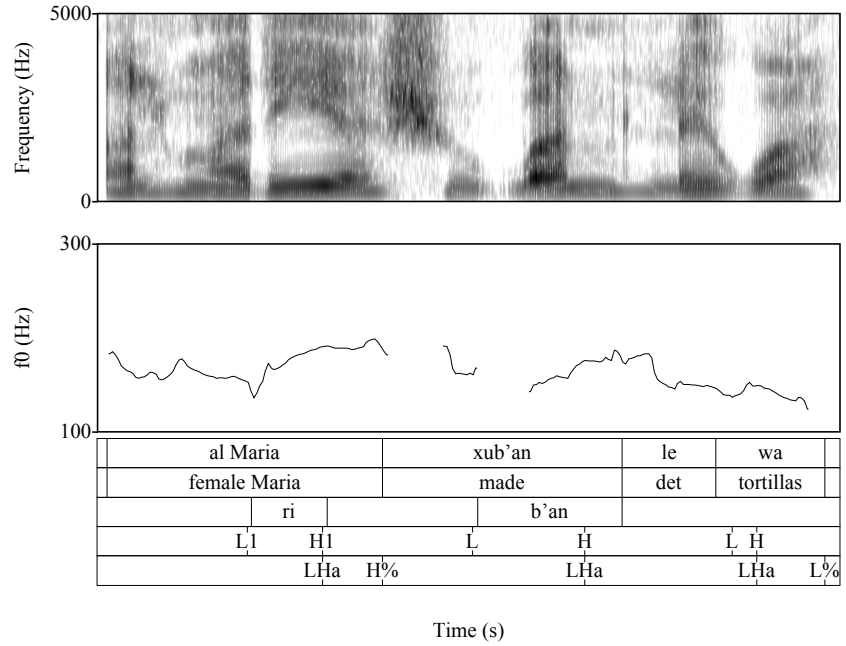

Figure 3: An example focus sentence

Following the previous work on the prosody of focus and contrastive topic that was discussed earlier, I have also looked at the following variables: (i) the duration of the stressed syllable in the pre-predicate expression, (ii) the alignment of the rising tone with respect to the onset of the stressed syllable, (iii) the duration of the rise of the F0 contour, (iv) the range of the rise, and (v) the slope of the rise. Figures 3 and 4 provide two example utterances illustrating how the analyses were carried out. In these figures, the first two tiers give the words and the glosses, respectively. The stressed syllables of the pre-predicate expression and the verb are marked in the third tier. The fourth tier provides information about the local minimum and maximum of the rises associated with each content word which are marked by the letters L and H, respectively. The last tier marks the rising pitch movement associated with each content word by LHa as well as the boundary tones, e.g. L% or H%.

Given these conventions about the annotation, Table 3 shows how the variables mentioned above are calculated. Here, I use t(x) to indicate the time corresponding to x

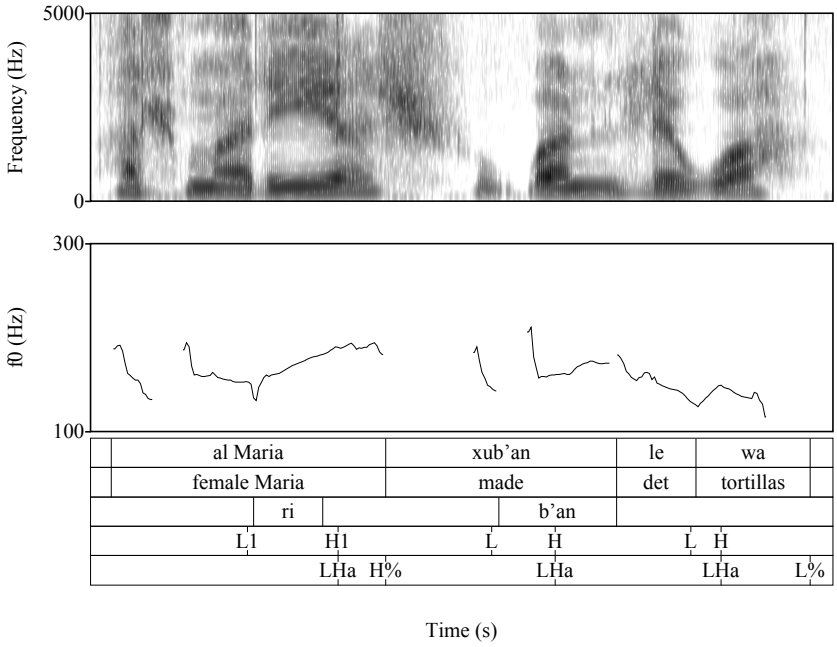

Figure 4: An example contrastive topic sentence

and $f_0(x)$ to indicate the F0 value corresponding to x:

Alignment of the L tone	t(L)-t(onset of the stressed syllable)
Alignment of the H tone	t(H)-t(onset of the stressed syllable)
Duration of the F0 rise	t(H)-t(L) (=D)
Range of the F0 rise	f_0(H)-f_0(L) (=R)
Slope of the F0 rise	R/D

Table 3: Calculation of the phonetic variables

Each dependent variable was fitted in a linear mixed effects model with speaker and item as random variables and condition as an independent variable. Among the measurements that were taken, there was a ∼6Hz difference in the range of the F0 rise across conditions and the linear models yielded a statistically significant result ($p < 0.05$) only for this variable. Figure 5 is a box plot that shows the distribution of the range of the rise on the pre-predicate expression across conditions. Given that this difference is very small, it may or may not be perceivable by K'iche' listeners. A perception study is needed to find out whether such a small difference is indeed perceivable.

I now turn to a discussion of the prosody of the post-focal or post-topical parts of the target sentences. A total of 104 utterances (64%) had a rising pitch movement on the verb. For these verbs, I carried out the same measurements as above. The remaining

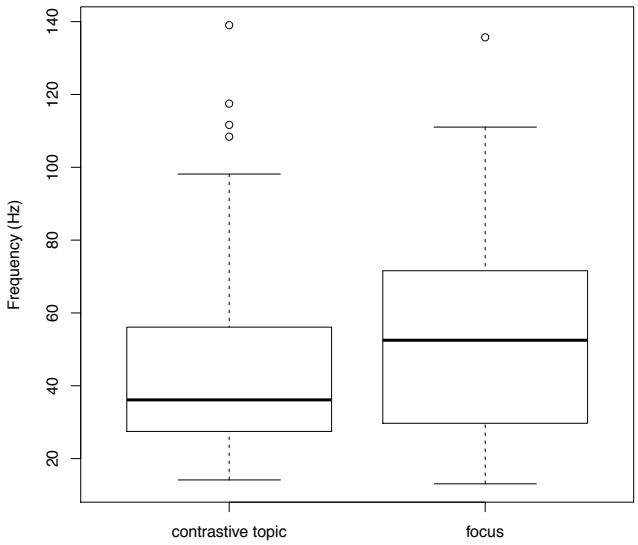

Figure 5: Range of the F0 rise by conditions

62 utterances did not have a rise on the verb either because (i) there was tonal truncation (n=54), or (ii) the F0 was flat on the verb (n=2), or (iii) the L or the H target could not be identified (n=6). For these cases, I only compared the H tone realized on the verb (if at all) across conditions. For each verb, I also looked at the duration of the stressed syllable.

As above, each dependent variable was fitted in a linear mixed effects model with speaker and item as random variables and condition as an independent variable. The linear mixed effects models did not yield significant effects for any of the variables.

5 Discussion

In the prosodic analyses of the data, each target sentence was divided into two parts, namely, a pre-predicate part and a post-focal or post-topical part, to be able to clearly see the predictions of the previous literature. Contrary to what Can Pixabaj & England (2011) claim, there is no clear indication that focus and contrastive topic are distinguished by a pause between the pre-predicate expression and the post-focal/post-topical expression. On the other hand, the data showed that there is a rising pitch movement associated with the focused or contrastively topicalized expression and a linear mixed effects model yielded a significant effect of condition on the range of this rise. However, the difference across conditions was small ($\sim$6 Hz) and requires a perception study to determine if such a difference matters for listeners. If the small difference in the F0 range that turned out to be significant in this study is actually not perceivable, then context may be the only source of

the intended interpretation.

In general, one can assume that pragmatic meanings are reflected in prosody because they are represented in the speaker's cognitive model and figure into speech planning. However, at a theoretical level, it is also possible that such an effect of pragmatic meanings on prosody may not always exist. Indeed, there are a series of studies on Yukatek Maya, e.g. Kügler *et al.* (2007); Kügler & Skopeteas (2006), which claim that there is no interaction between topic/focus and pitch manipulations. More generally, recent work suggests that there are languages where no prosodic reflexes of information structure are observed, e.g. Northern Sotho (Zerbian 2006), Hausa (Hartmann & Zimmermann 2007), Wolof (Rialland & Robert 2001) and Thompson River Salish (Koch 2008). A commonality across these languages is the use of word order changes and/or morphology to indicate the changes in information structure. The K'iche' data show that something similar might be going on in K'iche', especially if the significant difference in the range of the F0 rise is not perceivable by listeners.

6 Conclusion

This paper presented an experimental study on K'iche' designed to identify whether string-identical sentences with either a focus or a contrastive topic interpretation differ in their prosodic properties. The experiment involved obtaining naturally occurring data from native speakers of K'iche' by having them repeat target sentences they heard in conversations. The acoustic analyses of several variables yielded a significant effect of condition only in the range of the F0 rise associated with focused and contrastively topicalized expressions. However, the difference across conditions is only $\sim$6 Hz which may not be perceivable by listeners. Contrary to previous studies, the data did not support the claim that existence of a pause following the pre-predicate expressions distinguishes contrastive topics from foci.

The future research for this project can proceed in two directions. One is to improve the current experiment by using nominals in the pre-predicate position to prevent any interference of Spanish. The new stimuli should also include non-contrastively topicalized expressions in the pre-predicate position to determine the prosodic properties of such expressions and to compare them with the other two. Lastly, the experiment should be run on more participants in order to obtain a more representative sample. The second direction is to design a perception experiment where the data from the production experiment are used as stimuli. The perception experiment can be designed so that a given target utterance, say one where the pre-predicate expression is focused, can occur both in focus and contrastive topic contexts and the listeners can be asked to judge the acceptability of such utterances. If participants consistently accept a given target utterance in either context, then this shows that they do not make a prosodic distinction between focus and contrastive topic. Results from such an experiment would prove to be useful in interpreting the results of a production experiment especially if statistically significant but phonetically small differences are found.

References

AISSEN, JUDITH. 1992. Topic and focus in Mayan. *Language* 68.43–80.

——. 1999. Agent focus and inverse in Tzotzil. *Language* 75.451–485.

——. 2011. On the syntax of agent focus in K'ichee'. In *Proceedings of FAMLi I*, ed. by Kirill Shklovsky, Pedro Mateo Pedro, & Jessica Coon. MIT Working Papers in Linguistics.

BRAUN, BETTINA. 2006. Phonetics and phonology of thematic contrast in German. *Language and Speech* 49.451–493.

BÜRING, DANIEL. 2003. On d-trees, beans, and b-accents. *Linguistics and Philosophy* 5.511–545.

——. 2011. Focus. In *The Cambridge Encyclopedia of the Language Sciences*, ed. by Patrick Colm Hogan. Cambridge: Cambridge University Press.

CAN PIXABAJ, TELMA. 2010. La predicación secundaria en K'ichee': Una construcción restringada. In *La predicación secundaria en lenguas de mesoamérica*, ed. by Judith Aissen & Roberto Zavala. Mexico City: CIESAS.

——, & NORA ENGLAND. 2011. Nominal topic and focus in K'ichee'. In *Representing Language: Essays in Honor of Judith Aissen*, ed. by Rodrigo Guitérrez-Bravo, Line Mikkelsen, & Eric Potsdam, 15–30. Linguistics Research Center, UC Santa Cruz Department of Linguistics.

COOPER, WILLIAM E., STEPHEN J. EADY, & PAMELA R. MUELLER. 1985. Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America* 77.21–42.

DAYLEY, JON P. 1981. Voice and ergativity in Mayan languages. *Journal of Mayan Linguistics* 2.3–82.

DIXON, R.M.W. 1994. *Ergativity*. Cambridge: Cambridge University Press.

EADY, STEPHEN J., WILLIAM E. COOPER, GAYLE V. KLOUDA, PAMELA R. MUELLER, & DAN W. LOTTS. 1986. Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech* 29.233–251.

ENGLAND, NORA. 1991. Changes in basic word order in Mayan languages. *International Journal of American Linguistics* 57.446–486.

FACE, TIMOTHY. 2001. Focus and early peak alignment in Spanish intonation. *Probus* 13.223–246.

——. 2002. Local intonational marking of Spanish contrastive focus. *Probus* 14.71–92.

FÉRY, CAROLINE. 2006. The prosody of topicalization. In *On information structure, meaning and form: Generalizations across languages*, ed. by Kerstin Schwabe & Susanne Winkler, 69–86. Amsterdam, John Benjamins.

——, & VIERI SAMEK-LADOVICI. 2006. Focus projection and prosodic prominence in nested foci. *Language* 82.131–150.

GRICE, MARTINE, MARIAPAOLA D'IMPERIO, MICHELINA SAVINO, & CINZIA AVESANI. 2005. Strategies for intonation labelling across varieties of Italian. In *Prosodic typology*, ed. by S.-A. Jun, 362–289. Oxford: Oxford University Press.

HARTMANN, KATHARINA, & MALTE ZIMMERMANN. 2007. In place – out of place: Focus in Hausa. In *On information structure, meaning and form: Generalizations across languages*, ed. by Kerstin Schwabe & Susanne Winkler, 365–403. Amsterdam: John Benjamins.

HENDERSON, ROBERT. 2012. Morphological alternations at the intonational phrase edge. *Natural Language and Linguistic Theory* 30.741–787.

JACKENDOFF, RAY S. 1972. *Semantic interpretation in Generative Grammar*. Cambridge, MA: The MIT Press.

JUN, SUN-AH. 2005. Prosodic typology. In *Prosodic Typology: The Phonology of Intonation and Phrasing*, ed. by S.-A. Jun. Oxford: Oxford University Press.

KADMON, NIRIT. 2001. *Formal pragmatics*. Malden, MA: Blackwell.

KOCH, KARSTEN. 2008. *Intonation and focus in Nte?kepmxcin (Thompson River Salish)*. University of British Columbia dissertation.

KÜGLER, FRANK, & STAVROS SKOPETEAS. 2006. Interaction of lexical tone and information structure in Yucatec Maya. In *International Symposium on Tonal Aspects of Language*, 83–88, La Rochelle.

——, STAVROS SKOPETEAS, & ELISABETH VERHOEVEN. 2007. Encoding information structure in Yucatec Maya: on the interplay of prosody and syntax. *Interdisciplinary Studies on Information Structure* 8.187–208.

LARSEN, THOMAS W. 1987. The syntactic status of ergativity in Quiché. *Lingua* 71.33–59.

—— 1988. *Manifestations of ergativity in Quiché grammar*. University of California, Berkeley dissertation.

LÓPEZ IXCOY, CANDELARIA DOMINGA. 1997. *Ri ukemiik ri K'ichee' chii': Gramática K'ichee'*. Guatemala City: Cholsamaj.

MONDLOCH, JAMES L. 1978. *Basic Quiché grammar*. Institute for Mesoamerican Studies, State University of New York at Albany.

—— 1981. *Voice in Quiché-Maya*. State University of New York at Albany dissertation.

NIELSEN, KUNIKO. 2005. K'iche' intonation. *UCLA Working Papers in Phonetics* 104.45–60.

NORMAN, WILLIAM. 1977. Topic and focus in Mayan. In *Presentation at the Mayan Workshop II*, San Cristóbal de las casas, Chiapas, México.

PIERREHUMBERT, JANET. 1980. *The phonology and phonetics of English intonation*. MIT dissertation.

PYE, CLIFTON. 2001. The acquisition of finiteness in K'iche' Maya. In *Proceedings of the 25th Annual Boston University Conference on Language Development*, 645–656.

——, & PEDRO QUIXTAN POZ. 1988. Precocious passives (and antipassives) in Quiché Mayan. *Papers and reports on child language development* 27.71–80.

RIALLAND, ANNIE, & STÉPHANE ROBERT. 2001. The intonational system of Wolof. *Linguistics* 39.893–939.

RICHARDS, MICHAEL. 2003. *Atlas Lingüistico de Guatemala*. Guatemala City: Universidad de Rafael Landívar.

ROBERTS, CRAIGE. 1996. Information structure: Towards an integrated formal theory of pragmatics. In *OSUWPL*, ed. by Jae Hak Yoon & Andreas Kathol, volume 49, 91–136. The Ohio State University, Department of Linguistics. Reprinted with a new Afterword in *Semantics and Pragmatics* volume 5, 2012.

——. 1998. Focus, the flow of information, and Universal Grammar. In *The Limits of Syntax*, ed. by Peter Culicover & Louise McNally, 109–160. New York: Academic Press.

——. 2012. Topics. In *Semantics: An International Handbook of Natural Language Meaning*, ed. by Claudia Maienborn, Klaus von Heusinger, & Paul Portner, volume 33.2, 1908–1934. Mouton de Gruyter.

ROMERO, SERGIO. 2012. A Maya version of Jespersen's Cycle: The diachronic evolution of negative markers in K'iche' Maya. *International Journal of American Linguistics* 78.77–96.

ROOTH, MATS. 1992. A theory of focus interpretation. *Natural Language Semantics* 1.75–116.

——. 1996. Focus. In *The Handbook of Contemporary Semantic Theory*, ed. by Shalom Lappin. London: Blackwell.

STIEBELS, BARBARA. 2006. Agent focus in Mayan languages. *Natural Language and Linguistic Theory* 24.501–570.

TRECHSEL, FRANK R. 1993. Quiché focus constructions. *Lingua* 91.33–78.

VENDITTI, JENNIFER J., KIKUO MAEKAWA, & MARY BECKMAN. 2008. Prominence marking in the Japanese intonation system. In *Handbook of Japanese Linguistics*, ed. by Shigeru Miyagawa & Mamuro Saito, 456–512. Oxford: Oxford University Press.

WANG, BEI, & XI YU. 2011. Differential prosodic encoding of topic and focus in sentence-initial position in Mandarin Chinese. *Journal of Phonetics* 39.595–611.

ZERBIAN, SABINE. 2006. *Expression of information structure in the Bantu language Northern Sotho*. Humboldt University dissertation.