Marie-Catherine de Marneffe demarneffe.1@osu.edu

Research statement - October 2016 http://www.ling.ohio-state.edu/∼demarneffe.1/

One of my research goals is to develop computational linguistic methods that capture what is conveyed by language beyond the literal meaning of the words. When faced with a piece of text, humans understand far more than just the literal meaning of the words in the text. In even our most mundane interactions, much of what we communicate is not said explicitly but rather inferred. My dissertation work already highlighted the necessity of capturing, in natural language processing (NLP), the inferences that we draw from text, and illustrated the approach I take to solving that challenge. I leverage crowdsourcing techniques to tap into people's intuitions to obtain reliable judgements on what people infer from text, and construct supervised models based on simple but efficient machine learning techniques in which feature engineering is driven by deep linguistic and error analysis.

My work also aims at demonstrating how linguistically-motivated features, conjoined with surface-level ones, are necessary to enable fundamental progress in achieving robust text understanding. Research in NLP has been moving into the direction of learning from big data, downsizing the necessity of linguistic theory as a firm basis for automatic language processing: sufficient surface-level features (i.e., n-grams) should be capable of approximating simultaneously linguistic constraints present in the data and usage patterns. Statistical methods exploiting surface features have indeed led us quite far in building language processing technologies. But they are reaching a limit: for instance, despite the massive amount of data Google possesses, its phrase-based machine translation is not reaching human-level performance. My work shows how linguistic features can contribute to understanding, when combined with surface-level features. I have shown that, for coreference resolution, linguistic features work best when combined with a larger set of surface features, giving an additional gain in performance than when using surface features alone (de Marneffe, Recasens & Potts 2015). Recently, as part of my NSF CRII grant, I investigated the identification of agreement/disagreement in online debate posts, where agreement or disagreement is often implicitly expressed, and showed that by leveraging the progress made in monolingual alignment and using semantic features (such as polarity mismatch between sentences as well as embeddings under modality or negation) on top of surface-level features, we improve the recall of disagreements (Gokcen & de Marneffe 2015).

I am in the process of expanding the PragBank corpus I developed (de Marneffe, Potts & Manning 2012) which aims at studying how hearers/readers assess the veridicality of events. The PragBank corpus is small (642 sentences), it is restricted to newswire data, and sentences were presented in isolation, so there was no interaction of discourse. The corpus will now contain short passages (instead of sentences in isolation) from newswire, fiction, spoken and social media genres. This extended corpus will allow us to explore systematically the factors playing a role in veridicality interpretation from a theoretical linguistic perspective, by applying machine learning techniques

(such as feature selection techniques) to corpus data. Second, building on the linguistic analysis to define features, the corpus will serve as training data to develop a veridicality classifier, not restricted to the newswire genre, that will have practical implications for NLP tasks, such as information extraction and textual entailment.

One of the factors that play a role in which inferences are drawn from utterances is prosody. With Judith Tonhauser (OSU Linguistics), I investigated experimentally the impact of prosody in drawing inferences from indirect answers to polar questions (specifically the difference between a neutral contour and the rise-fall-rise contour). We showed that the prosodic realization of an indirect answer can provide a cue to the speaker's intended meaning, and discussed implications of our findings for scalar implicature generation and the meaning of the rise-fall-rise contour (de Marneffe & Tonhauser, to appear).

I have also been one of the principal developers of the Universal Dependencies project, which seeks to build cross-linguistically consistent treebank annotation for many languages, with the goal of facilitating multilingual parser development, cross-lingual learning, and parsing research from a language typology perspective (`http://universaldependencies.github.io/docs/`). It is based on the (universal) Stanford dependencies (de Marneffe et al. 2006, 2008, 2014), Google universal part-of-speech tags (Petrov et al. 2012), and the Interset interlingua for morphosyntactic tagsets (Zeman 2008). Recently, I wrote a paper which outlines the goals of the project, gives the principles for the annotation, and describes the datasets released so far (Nivre, de Marneffe et al. 2016: Joakim Nivre is the first author to reflect that he is the leader of the project).

At the intersection of my work on veridicality and linguistic annotation, Micha Elsner (OSU Linguistics) and I built a corpus of child-directed utterances involving mental state verbs (e.g., *think*, *know*, *remember*), annotated for utterance purpose (does the utterance report a belief, or mainly asserts something) and veridicality (is the complement of the mental state verb true or false). Our work provides evidence for the pragmatic hypothesis of Lewis et al. (2012) on why children have difficulty with false beliefs (de Marneffe & Elsner 2015; de Marneffe & Elsner, in preparation). In collaboration with Shari Speer (OSU Linguistics) and Laura Wagner (OSU Psychology), we are now investigating experimentally at what age children become sensitive to prosody and pragmatic factors which tease apart the uses of the verb *think* as belief reports or polite assertions.