The Interaction of Syntactic Theory and Computational Psycholinguistics

Frank Keller

School of Informatics, University of Edinburgh
10 Crichton Street, Edinburgh EH8 9AB, UK
`keller@inf.ed.ac.uk`

1 Introduction

Typically, current research in psycholinguistics does not rely heavily on results from theoretical linguistics. In particular, most experimental work studying human sentence processing makes very straightforward assumptions about sentence structure; essentially only a simple context-free grammar is assumed. The main text book in psycholinguistics, for instance, mentions Minimalism in its chapter on linguistic description (Harley, 2001, ch. 2), but does not provide any details, and all the examples in this chapter, as well as in the chapters on sentence processing and language production (Harley, 2001, chs. 9, 12), only use context-free syntactic structures with uncontroversial phrase markers (S, VP, NP, etc.). The one exception is traces, which the textbook discusses in the context of syntactic ambiguity resolution.

Harley's (2001) textbook is typical of experimental psycholinguistics, a field in which most of the work is representationally agnostic, i.e., assumptions about syntactic structure left implicit, or limited to uncontroversial ones. However, the situation is different in computational psycholinguistics, where researchers built computationally implemented models of human language processing. This typically involves making one's theoretical assumptions explicit, a prerequisite for being able to implement them. For example, Crocker's (1996) model explicitly implements assumptions from the Principles and Parameters framework, while Hale (2006) uses probabilistic Minimalist grammars, or Mazzei et al. (2007) tree-adjoining grammars.

Here, we will investigate how evidence regarding human sentence processing can inform our assumptions about syntactic structure, at least in so far as this structure is used in computational models of human parsing.

2 Prediction and Incrementality

Evidence from psycholinguistic research suggests that language comprehension is largely *incremental*, i.e., that comprehenders build an interpretation of a sentence on a word-by-word basis. Evidence for incrementality comes from speech shadowing, self-paced reading, and eye-tracking studies (Marslen-Wilson, 1973; Konieczny, 2000; Tanenhaus et al., 1995): as soon as a reader or listener perceives a word in a sentence, they integrate it as fully as possible into a representation of the sentence thus far. They experience differential processing difficulty during this integration process, depending on the properties of the word and its relationship to the preceding context.

There is also evidence for full *connectivity* in human language processing (Sturt and Lombardo, 2005). Full connectivity means that all words are connected by a single syntactic structure; the parser builds no unconnected tree fragments, even for the incomplete sentences (sentence prefixes) that arise during incremental processing.

Furthermore, there is evidence that readers or listeners make *predictions* about upcoming material on the basis of sentence prefixes. Listeners can predict an upcoming post-verbal element, depending on the semantics of the preceding verb (Kamide et al., 2003). Prediction effects can also be observed in reading. Staub and Clifton (2006) showed that following the word *either* readers predict the conjunction *or* and the complement that follows it; processing was facilitated compared to structures that include *or* without *either*. In an ERP study, van Berkum et al. (1999) found that listeners use contextual information to predict specific lexical items and experience processing difficulty if the input is incompatible with the prediction.

The concepts of incrementality, connectedness, and prediction are closely related: in order to guar-

antee that the syntactic structure of a sentence prefix is fully connected, it may be necessary to build phrases whose lexical anchors (the words that they relate to) have not been encountered yet. Full connectedness is required to ensure that a fully interpretable structure is available at any point during incremental sentence processing.

Here, we explore how these key psycholinguistic concepts (incrementality, connectedness, and prediction) can be realized within a new version of tree-adjoining grammar, which we call Psycholinguistically Motivated TAG (PLTAG). We propose a formalization of PLTAG and a linking theory that derives predictions of processing difficulty from it. We then present an implementation of this model and evaluate it against key experimental data relating to incrementality and prediction. This approach is described in more detail in Demberg and Keller (2008) and Demberg and Keller (2009).

3 Modeling Explicit Prediction

We propose a theory of sentence processing guided by the principles of incrementality, connectedness, and prediction. The core assumption of our proposal is that a sentence processor that maintains explicit predictions about the upcoming structure has to validate these predictions against the input it encounters. Using this assumption, we can naturally combine the forward-looking aspect of Hale's (2001) surprisal (sentence structures are computed incrementally and unexpected continuations cause difficulty) with the backward-looking integration view of Gibson's (1998) dependency locality theory (previously predicted structures are verified against new evidence, leading to processing difficulty as predictions decay with time).

In order to build a model that implements this theory, we require an incremental parser that is capable of building fully connected structures and generating explicit predictions from which we can then derive a measure of processing difficulty. Existing parsers and grammar formalism do not meet this specification. While there is substantial previous work on incremental parsing, none of the existing model observes full connectivity. One likely reason for this is that full connectivity cannot be achieved using canonical linguistic structures as assumed in standard grammar formalisms such as CFG, CCG, TAG, LFG, or HPSG. Instead, a stack has to be used to store partial structures and retrieve them later when it has become clear

(through additional input) how to combine them.

We therefore propose a new variant of the tree-adjoining grammar (TAG) formalism which realizes full connectedness. The key idea is that in cases where new input cannot be combined immediately with the existing structure, we need to predict additional syntactic material, which needs to be verified against future input later on.

4 Incremental Processing with PLTAG

PLTAG extends normal LTAG in that it specifies not only the canonical lexicon containing lexicalized initial and auxiliary trees, but also a predictive lexicon which contains potentially unlexicalized trees, which we will call *prediction trees*. Each node in a prediction tree is annotated with indices of the form $_{s_j}^{s_j}$, where inner nodes have two identical indices, root nodes only have a lower index and foot and substitution nodes only have an upper index. The reason for only having half of the indices is that these nodes (root, foot, and substitution nodes) still need to combine with another tree in order to build a full node. If an initial tree substitutes into a substitution node, the node where they are integrated becomes a full node, with the upper half contributed by the substitution node and the lower half contributed by the root node.

Prediction trees have the same shape as trees from the normal lexicon, with the difference that they do not contain substitution nodes to the right of their spine (the spine is the path from the root node to the anchor), and that their spine does not have to end with a lexical item. The reason for the missing right side of the spine and the missing lexical item are considerations regarding the granularity of prediction. This way, for example, we avoid predicting verbs with specific subcategorization frames (or even a specific verb). In general, we only predict upcoming structure as far as we need it, i.e., as required by connectivity or subcategorization. (However, this is a preliminary assumption, the optimal prediction grain size remains an open research question.)

PLTAG allows the same basic operations (substitution and adjunction) as normal LTAG, the only difference is that these operations can also be applied to prediction trees. In addition, we assume a verification operation, which is needed to validate previously integrated prediction trees. The tree against which verification happens has to always match the predicted tree in shape (i.e., the

verification tree must contain all the nodes with a unique, identical index that were present in the prediction tree, and in the same order; any additional nodes present in the verification tree must be below the prediction tree anchor or to the right of its spine). This means that the verification operation does not introduce any tree configurations that would not be allowed by normal LTAG. Note that substitution or adjunction with a predictive tree and the verification of that tree always occur pairwise, since each predicted node has to be verified. A valid parse for a sentence must not contain any nodes that are still annotated as being predictive – all of them have to be validated through verification by the end of the sentence.

5 Modeling Processing Difficulty

Our variant of TAG is designed to implement a specific set of assumptions about human language processing (strong incrementality with full connectedness, prediction, ranked parallel processing). The formalism forms the basis for the processing theory, which uses the parser states to derive estimates of processing difficulty. In addition, we need a linking theory that specifies the mathematical relationship between parser states and processing difficulty in our model.

During processing, the elementary tree of each new word is integrated with any previous structure, and a set of syntactic expectations is generated (these expectations can be easily read off the generated tree in the form of predicted trees). Each of these predicted trees has a time-stamp that encodes when it was first predicted, or last activated (i.e., accessed). Based on the timestamp, a tree's decay at verification time is calculated, under the assumption that recently-accessed structures are easier to integrate.

In our model, processing difficulty is thus incurred during the construction of the syntactic analyses, as calculated from the probabilities of the elementary trees (this directly corresponds to Haleian surprisal calculated over PLTAG structures instead of over CFG structures). In addition to this, processing difficulty has a second component, the cost of verifying earlier predictions, which is subject to decay.

The verification cost component bears similarities to DLT integration costs, but we do not calculate distance in terms of number of discourse referents intervening between a dependent and its head.

Rather verification cost is determined by the number of words intervening between a prediction and its verification, subject to decay. This captures the intuition that a prediction becomes less and less useful the longer ago it was made, as it decays from memory with increasing distance.

6 Evaluation

In Demberg and Keller (2009), we present an implementation of the PLTAG model, including a lexicon induction procedure, a parsing algorithm, and a probability model. We show that the resulting framework can capture experimental results from the literature, and can explain both locality and prediction effects, which standard models of sentence processing like DLT and surprisal are unable to account for simultaneously.

Our model therefore constitutes a step towards a unified theory of human parsing that potentially captures a broad range of experimental findings. This work also demonstrates that (computational) psycholinguistics cannot afford to be representationally agnostic – a comprehensive, computationally realistic theory of human sentence processing needs to make explicit assumptions about syntactic structure. Here, we showed how the fact that human parsing is incremental and predictive necessitates certain assumptions about syntactic structure (such as full connectedness), which can be implemented by augmenting an existing grammar formalism, viz., tree-adjoining grammar. Note, however, that it is difficult to show that this approach is the only one that is able to realize the required representational assumptions; other solutions using different grammar formalisms are presumably possible.

7 Acknowledgments

References

Crocker, Matthew W. 1996. *Computational Psycholinguistics: An Interdisciplinary Approach to the Study of Language*. Kluwer, Dordrecht.

Demberg, Vera and Frank Keller. 2008. A psycholinguistically motivated version of TAG. In

Proceedings of the 9th International Workshop on Tree Adjoining Grammars and Related Formalisms. Tübingen.

Demberg, Vera and Frank Keller. 2009. A computational model of prediction in human parsing: Unifying locality and surprisal effects. Manuscript under review.

Gibson, Edward. 1998. Linguistic complexity: locality of syntactic dependencies. *Cognition* 68:1–76.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, Pittsburgh, PA, volume 2, pages 159–166.

Hale, John. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30(4):609–642.

Harley, Trevor. 2001. *The Psychology of Language: From Data to Theory*. Psychology Press,, Hove, 2 edition.

Kamide, Yuki, Christoph Scheepers, and Gerry T. M. Altmann. 2003. Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research* 32:37–55.

Konieczny, Lars. 2000. Locality and parsing complexity. *Journal of Psycholinguistic Research* 29(6):627–645.

Marslen-Wilson, William D. 1973. Linguistic structure and speech shadowing at very short latencies. *Nature* 244:522–523.

Mazzei, Alessandro, Vicenzo Lombardo, and Patrick Sturt. 2007. Dynamic TAG and lexical dependencies. *Research on Language and Computation* 5(3):309–332.

Staub, Adrian and Charles Clifton. 2006. Syntactic prediction in language comprehension: Evidence from either . . . or. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32:425–436.

Sturt, Patrick and Vincenzo Lombardo. 2005. Processing coordinated structures: Incrementality and connectedness. *Cognitive Science* 29(2):291–305.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268:1632–1634.

van Berkum, J. J. A., C. M. Brown, and Peter Hagoort. 1999. Early referential context effects in sentence processing: Evidence from event-related brain potentials. *Journal of Memory and Language* 41:147–182.