Challenging incrementality in human language processing: two operations for a cognitive architecture

Philippe Blache

Aix-Marseille Université & CNRS
LPL (UMR7309), 13100, Aix-en-Provence, France
blache@blri.fr

The description of language complexity and the cognitive load related to the different linguistic phenomena is a key issue for the understanding of language processing. Many studies have focused on the identification of specific parameters that can lead to a simplification or on the contrary to a complexification of the processing (e.g. the different difficulty models proposed in (Gibson, 2000), (Warren and Gibson, 2002), (Hawkins, 2001)). Similarly, different simplification factors can be identified, such as the notion of activation, relying on syntactic priming effects making it possible to predict (or activate) a word (Vasishth, 2003). Several studies have shown that complexity factors are cumulative (Keller, 2005), but can be offset by simplification (Blache et al., 2006). It is therefore necessary to adopt a global point of view of language processing, explaining the interplay between positive and negative cumulativity, in other words compensation effects.

From the computational point of view, some models can account more or less explicitly for these phenomena. This is the case of the Surprisal index (Hale, 2001), offering for each word an assessment of its integration costs into the syntactic structure. This evaluation is done starting from the probability of the possible solutions. On their side, symbolic approaches also provide an estimation of the activation degree, depending on the number and weight of syntactic relations to the current word (Blache et al., 2006); (Blache, 2013).

These approaches are based on the classical idea that language processing is incremental and occurs word by word. There are however several experimental evidences showing that a higher level of processing is used by human subjects. Eye-tracking data show for example that fixations are done by chunks, not by words (Rauzy and Blache, 2012). Similarly, EEG experiments have shown that processing multiword expressions (for example idioms) relies on global mechanisms (Vespignani et al., 2010); (Rommers et al., 2013).

Starting from the question of complexity and its estimation, I will address in this presentation the problem of language processing and its organization. I propose more precisely, using computational complexity models, to define a cohesion index between words. Such an index makes it possible to define chunks (or more generally units) that are built directly, by aggregation, instead of syntactic analysis. In this hypothesis, parsing consists in two different processes: aggregation and integration.

Acknowledgments

Short biography

Philipe Blache is Senior Researcher at CNRS (Aix-Marseille University, France). He is the Director of the BLRI (*Brain and Language Research Institute*), federating 6 research laboratories in Linguistics, Computer Science, Psychology and Neurosciences.

Philippe Blache earned an MA in Linguistics from Université de Provence and a MSc in Computer Science from Université de la Méditerranée, where he received in 1990 his PhD in Artificial Intelligence.

During his career, Philippe Blache has focused on Natural Language Processing and Formal Linguistics, with a special interest in spoken language analysis. He has proposed a linguistic theory, called *Property Grammars*, suitable for describing language in its different uses, and explaining linguistic domains interaction. His current aca-

1

demic works address the question of human language processing and its complexity.

Philippe Blache has been director of two CNRS laboratories in France (2LC and LPL). He has served on numerous boards (European Chapter of the ACL, ESSLLI standing committee, CSLP, etc.). He is currently member of the Scientific Council of Aix-Marseille Université, member of the "Comité National de la Recherche Scientifique" in computer science and he chairs the TALN conference standing committee.

References

Philippe Blache, Barbara Hemforth, and Stéphane Rauzy. 2006. Acceptability prediction by means of grammaticality quantification. In *ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, July.

Philippe Blache. 2013. Chunks et activation : un modèle de facilitation du traitement linguistique. In *Proceedings of TALN-2014*.

Edward Gibson. 2000. The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. In Alec Marantz, Yasushi Miyashita, and Wayne O'Neil, editors, *Image, Language, Brain*, pages 95–126. Cambridge, Massachussetts, MIT Press.

John Hale. 2001. A probabilistic earley parser as a psycholinguistic model. In *Proceeding of 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, Pittsburgh, PA.

John Hawkins. 2001. Why are categories adjacent. *Journal of Linguistics*, 37.

Frank Keller. 2005. Linear Optimality Theory as a Model of Gradience in Grammar. In *Gradience in Grammar: Generative Perspectives*. Oxford University Press.

Stéphane Rauzy and Philippe Blache. 2012. Robustness and processing difficulty models. a pilot study for eye-tracking data on the french treebank. In *Proceedings of the 1st Eye-Tracking and NLP workshop*.

Joost Rommers, Antje S Meyer, Peter Praamstra, and Falk Huettig. 2013. Neuropsychologia. *Neuropsychologia*, 51(3):437–447, February.

Shravan Vasishth. 2003. Quantifying processing difficulty in human sentence parsing: The role of decay, activation, and similarity-based interference. In *Proceedings of the European Cognitive Science Conference 2003*.

Francesco Vespignani, Paolo Canal, Nicola Molinaro, Sergio Fonda, and Cristina Cacciari. 2010. Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22(8):1682–1700.

Tessa Warren and Ted Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition*, 85:79–112.