Implications of psycho-computational modelling for Morphological Theory

Marcello Ferro, Claudia Marzi, Vito Pirrelli
Institute for Computational Linguistics, National Research Council, Pisa, Italy

Goals

The workshop intends to offer an international forum for discussing interdisciplinary prospects of integration between advances in computer modelling of word knowledge and novel theoretical approaches to morphology.

Motivation and background

Following the advent of connectionism (Rumelhart & McClelland 1986), linguistic theory, cognitive models of human language processing and computer language modelling have increasingly been sharing research questions and goals. The rationale for this convergence was presciently epitomized in the early 1980s by Marr's (1980) hierarchy of levels of understanding of a complex processing system. Accordingly, linguistic theory approaches language issues mostly at Marrian level 1 ("what do speakers do when they use language?"), while cognitive psychology and computational linguistics are chiefly concerned with level 2 issues ("how do they use it?"), and neurosciences with level 3 questions ("how is such behavior implemented in the brain?"). Although David Marr originally introduced his hierarchy to emphasize that explanations at different levels can be investigated independently of each other, over the last 25 years there has been growing interest in the potential for between-level interaction, with a view to investigating the methodological conditions for their interdisciplinary unification. Advances in computer sciences and neuro-imaging technology have provided the level of material continuity between linguistic functions (level 1), algorithmic operations (level 2) and neuro-functional correlates (level 3) that is a necessary pre-condition to successful integration of neighboring disciplines along Marr's hierarchy (Alvargonzáles 2011).

This trend represents a challenge and an opportunity for Morphological Theory. Computer simulations can spawn novel explanatory paradigms. The idea that linguistic structure can emerge through self-organization of unstructured input is nowadays key to our understanding of language acquisition (Bybee & Hopper 2001;

Ellis & Larsen-Freeman 2006; MacWhinney 1999; MacWhinney & O'Grady 2015). Nonetheless, it had to await the challenging test of successful computer simulations before being given wide currency in the psycholinguistic (Baayen et al. 2011) and theoretical literature (Blevins 2016).

A recent reconceptualization of morphological generalization as the "Cell Filling Problem" (Ackerman & Malouf 2013) hinges on modelling the implicative structure of morphological paradigms through conditional entropy, an information-theoretic measure of inferential complexity that proves to correlate significantly with speakers' behavior (Ferro et al. 2018; Milin et al. 2009a, 2009b). The task is carried out successfully with either deep learning architectures (Malouf 2017; Cardillo et al. 2018) or linear mappings (Baayen et al. 2018), showing that multiple inferences from a set of paradigmatically-related forms can further reduce the complexity of inflection learning (Bonami & Beniamine 2017).

Time-honored approaches like analogy-based synchronic descriptions of language systems and historical accounts of language change got a new lease of life when analogical relations and their cognitive implications were successfully operationalized in the machine learning literature (Albright 2002, 2009; Albright & Hayes 2003; Daelemans & van den Bosch 2005; Keuleers et al. 2007; Pirrelli & Yvon 1999).

In addition, computer models prove to be instrumental in breaking traditional theoretical deadlocks. To illustrate, the categorical subdivision between regularly and irregularly inflected forms advocated by dual models of word processing (Pinker & Ullman 2002), as well as Hockett's (1954) distinction between Item-and-Arrangement and Item-and-Process approaches to morphology, both rest on the assumption that storage and processing are two independent functions of the human language faculty. This assumption, however, is challenged by integrative, connectionist models of short-term and long-term memories, implemented as two different temporal dynamics of the same underlying process (Marzi & Pirrelli 2015).

Statistical language modelling has recently been used to test competing theoretical frameworks on a quantitative basis. For example, statistical analyses and computer simulations of speakers' reaction times in visual word recognition challenged evidence of amorphous, holistic representations in the speakers' mental lexicon (Lignos & Gorman 2012; Oseki et al. 2019; Virpioja et al. 2018).

Advances in distributional semantics (Baroni & Lenci 2010; Padó & Lapata 2007) have thrown in sharp relief the role of lexical semantics in morphological processing, particularly for compounding and derivation (Marelli et al. 2017; Marelli & Baroni 2015; Günther & Marelli 2018), while helping draw a measurably graded distinction between derivation and inflection (Bonami & Paperno 2018).

The issues

This is the right time to take stock of the implications of current computational models of word processing for morphological theory. We hope that the range of issues raised by the workshop will advance our understanding of issues spanning the entire Marr's hierarchy, from theoretical aspects to neuro-functional ones. In particular, we invite authors to address and discuss the following questions:

What are the optimal representation units of human morphological competence and how are they acquired? What role do they play in the way speakers process and store words? Do speakers combine these units in a linear way, as in chaining Markov models, or rather structure them hierarchically, as suggested by the literature on sentence processing? Do they store them in their long-term lexical repository economically, or rather multiply them redundantly, as a function of their context and use? In addition, are these units represented as independent items, or are they mutually related as nodes in a network of paradigmatic relations? What is the contribution of lexical semantics to this picture, and what type of influence is exercised on lexical units by the communicative context where they are used referentially?

What is the status of the processes combining these units into larger units? Are they implemented by a single mechanism? Or should we rather hypothesize that more than one mechanism is in place? What evidence do we have of the anatomical and functional localization of different combinatorial mechanisms in the brain? And in what ways do their neural implementations differ? Given the mounting evidence that children learn words in chunks and that ready-made stretches of assorted words are committed to the long-term memory by speakers, what does this evidence tell us about the separation between Morphology and Syntax for language learning? Can computer modelling sharpen our current understanding of issues of morphological complexity and their impact on lexical acquisition? What is its potential for modelling language learning, contact and change in multilingual contexts?

References

Ackerman, F. & Malouf, R. (2013). Morphological organization: the low conditional entropy conjecture. *Language* 89, 429–464.

Alvargonzález, D. (2011). Multidisciplinarity, Interdisciplinarity, Transdisciplinarity, and the Sciences. *International Studies in the Philosophy of Science* 25(4), 387-403.

Albright, A. (2009). Modeling analogy as probabilistic grammar. In Blevins, James P. and Blevins, Juliette (eds.), *Analogy in Grammar*, 185-213. Oxford: Oxford University Press.

Albright, A. (2002). Islands of reliability for regular morphology: Evidence from Italian. *Language* 78, 684-709.

Albright, A. & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition* 90, 119–161.

Baayen, H., Chuang, Y.Y., & Blevins, J.P. (2018) Inflectional morphology with linear mappings. *The Mental Lexicon* 13(2), 230-268.

Baayen H., Milin, P., Đurđević, D.F., Hendrix, P. & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118(3), 438-481.

Baroni, M., & Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics* 36(4), 673-721.

Blevins, J.P. (2016). *Word and paradigm morphology*. Oxford: Oxford University Press.

Bonami, O. & Paperno, D. (2018). Inflection vs. derivation in a distributional vector space, *Lingue e Linguaggio* XVII(2), 173-196.

Bybee, J. & Hopper, P.J. (eds.) (2001). *Frequency and the Emergence of Linguistic Structure*. John Benjamin Publishing Company.

Cardillo, A., Ferro, M., Marzi, M. & Vito Pirrelli (2018). Deep Learning of Inflection and the Cell-Filling Problem. *Italian Journal of Computational Linguistics* 4, 57-77.

Daelemans, W. and van de Bosch, A. (2005). *Memory-based Language Processing*. Cambridge, UK: Cambridge University Press.

Ellis, N.C. & Larsen-Freeman D. (2006). Language emergence: Implications for applied linguistics - Introduction to the special issue. *Applied Linguistics* 27, 558-589.

Ferro, M., Marzi, C. & Pirrelli, V. (2018). Discriminative word learning is sensitive to inflectional entropy. *Lingue e Linguaggio* XVII(2), 307-327.

Günther, F. & Marelli, M. (2018). Enter Sandman: Compound Processing and Semantic Transparency in a Compositional Perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition* http://dx.doi.org/10.1037/xlm0000677.

Hockett, C.F. (1954). Two models of grammatical description. *Word* 10, 210-231.

Keuleers, E., Sandra, D., Daelemans, W., Gillis, S., Durieux, G. & Martens, E. (2007). Dutch plural inflection: The exception that proves the analogy. *Cognitive Psychology*, 54, 283–318.

Lignos, C., & Gorman, K. (2012). Revisiting frequency and storage in morphological processing. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 48(1), 447-461. Chicago Linguistic Society.

MacWhinney, B. (ed.). (1999). *The emergence of language*. Lawrence Erlbaum Associates.

MacWhinney, B. & O'Grady, W. (eds.) (2015). *The Handbook of Language Emergence*. Wiley Blackwell.

Malouf, Robert (2017). Abstractive morphological learning with a recurrent neural network. *Morphology* 27(4), 431–458.

Marelli, M., Gagné, C. & Spalding, T. (2017). Compounding as Abstract Operation in Semantic Space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition* 166, 207-224.

Marelli, M., & Baroni, M. (2015). Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological review* 122(3), 485.

Marr, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Henry Holt, New York.

Marzi, C. & Pirrelli, V. (2015). A neuro-computational approach to understanding the mental lexicon. *Journal of Cognitive Science* 16(4): 493-534.

Milin, P., Đurđević, D.F., & del Prado Martín, F.M. (2009a). The simultaneous effects of inflectional paradigms and classes on lexical recognition: Evidence from Serbian. *Journal of Memory and Language* 60(1), 50-64.

Milin, P., Kuperman, V., Kostić, A., & Baayen, R. H. (2009b). Words and paradigms bit by bit: An information theoretic approach to the processing of paradigmatic structure in inflection and derivation. In J.P. Blevins & Blevins, J. (Eds.) *Analogy in grammar: Form and acquisition*, 214-252. Oxford University Press.

Oseki, Y., Yang, C. & Marantz, A. 2019. Modeling Hierarchical Syntactic Structures in Morphological Processing. In *Proceedings of Cognitive Modeling and Computational Linguistics* (CMCL) 2019, Minneapolis.

Padó, S. & Lapata, M. (2007). Dependency-based construction of semantic space models. *Computational Linguistics* 33 (2): 161–199

Pirrelli, V. & Yvon, F. (1999). The hidden dimension: a paradigmatic view of data-driven NLP. *Journal of Experimental and Theoretical Artificial Intelligence* 11(3), 391-408.

Virpioja, S., Lehtonen, M., Hultén, A., Kivikari, H., Salmelin, R. & Lagus, K. (2018). Using statistical models of morphology in the search for optimal units of representation in the human mental lexicon. *Cognitive science* 42(3), 939-973.