Prediction and Generalization in Word Processing and Storage

Word storage and processing have traditionally been modelled according to different computational paradigms, in line with the classical corner-stone of “dual-route” models of word structure assuming a sharp dissociation between memory and computation [1-5]. Even the most radical alternative to dual-route thinking, connectionist one-route models, challenged the lexicon-grammar dualism only by providing a neurally-inspired mirror image of classical base-to-inflection rules, while largely neglecting issues of lexical storage [6-8]. Recent psycho- and neuro-linguistic evidence, however, supports a less deterministic and modular view of the interaction between stored word knowledge and on-line processing [9-15]. The view entails simultaneous activation of distributed patterns of cortical connectivity encoding redundant distributional regularities in language data. Furthermore, recent developments in morphological theorizing question the primacy of grammar rules over lexical storage, arguing that word regularities emerge from independent principles of lexical organization, whereby lexical units and constructions are redundantly stored and mutually related through entailment relations [16-21]. We endorse here such a non modular view on Morphology to investigate two basic behavioural aspects of human word processing, morphological prediction and generalization, and offer a computer model supporting the hypothesis that they both derive from a common pool of principles of lexical organization.

Morphological generalization is at the roots of the human ability to develop expectations about novel word forms. Morphological prediction, on the other hand, defines the human capacity of anticipating upcoming known words. From the standpoint of memory processes, the two aspects are apparently in competition [22-24]. Prediction presupposes long-term entrenchment of dedicated Hebbian connections between time-bound receptors as the result of repeated exposure to frequent sequences of letters/segments. Long-term entrenchment eventually drives word recognition through anticipatory activation of frequently-activated Hebbian chains. Prediction is thus most accurate when concurrent activation chains are minimized. Computationally, this is equivalent to minimizing entropy over lexical choices. Generalization, on the other hand, requires that unseen chains of connections develop through learning, for novel words to be recognized as well-formed. This is equivalent to keeping entropy high in the lexicon to make room for unexperienced word stimuli.

We investigated the interplay between prediction and generalization in the lexicon by training two Artificial Topological Temporal Memories [25-26] on German and Italian text excerpts of about 5,000 word tokens each, sampled from corpora of children’s readings. To simulate low-level memory processes for serial order and their impact on morphological organization, only information about raw forms was given in training. After training, we probed the content of the two memories by tapping both seen and unseen word forms. The performance was measured in terms of recall accuracy and plotted against i) word form token frequency, ii) type/token frequency of the inflection paradigm and iii) type/token frequency of the paradigm cell. Simulation results prompt an interesting reappraisal of the role of morphological structure in word storage and processing. It is commonly assumed that structure-driven generalizations take centre-stage in language learning, and play the role of default mechanisms in language processing. In this perspective, the lexicon appears to play the subsidiary role of a fall-back, costly store of exceptions, a comparatively minor portion of morphological competence. In the light of the evidence reviewed here, this pattern is in fact reversed. Word processing and learning are primarily memory-driven processes, as pre-compilation of long memory “chains” is beneficial for prediction during on-line word processing. Morpheme-based generalizations are like shorter “chains” that come into the picture when memory of longer chains (whole words/constructions) fails, due to either novel, degenerate and noisy input, or to limitations in perception/memory spans. This explains the remarkably conservative nature of language learning, where over-regularization and levelling effects take place occasionally, and supports a more dynamic and less modularized view of language processing, where memory and computation, holistic and combinatorial knowledge, are in fact two sides of the same coin.

References

[1] Clahsen, H. (1999), Lexical entries and rules of language: A multidisciplinary study of German inflection, *Behavioral And Brain Sciences*, 22, 991–1060.

[2] Di Sciullo, A. M. and Williams, E. (1987), *On the Definition of Word*, Cambridge, MA: MIT Press.

[3] Pinker, S., and Prince, A. (1988), *On language and connectionism: Analysis of a parallel distributed processing model of language acquisition*, *Cognition*, 29, 195-247.

[4] Pinker, S., and Ullman, M.T. (2002), *The past and future of the past tense*, *Trends in Cognitive Science*, 6, 456-463.

[5] Prasada, S., and Pinker, S. (1993), Generalization of regular and irregular morphological patterns, *Language and Cognitive Processes* 8, 1-56.

[6] Rumelhart, D. E., and McClelland, J. L. (1986), *On learning of past tenses of English verbs*, in J.L. McClelland and D.E. Rumelhart (eds.), *Parallel distributed processing*, Cambridge, MA, MIT Press, 2, 216-271.

[7] McClelland, J., Patterson, K. (2002) Rules or connections in past-tense inflections: what does the evidence rule out? *Trends in Cognitive Science,* 6, 465-472.

[8] Seidenberg, M.S., and McClelland, J.L. (1989), *A distributed, developmental model of word recognition and naming*, in A. Galaburda (ed.), *From neurons to reading*, MIT Press.

[9] Baayen, H., Dijkstra, T., and Schreuder, R. (1997), Singulars and plurals in Dutch: Evidence for a parallel dual route model, *Journal of Memory and Language*, 37, 94-117.

[10] Hay, J. (2001), Lexical frequency in morphology: is everything relative?, *Linguistics*, 39, 1041-1111.

[11] Maratsos, M. (2000), More overregularizations after all, *Journal of Child Language*, 28, 32-54.

[12] Stemberger, J.P., and Middleton C.S. (2003), Vowel dominance and morphological processing, *Language and Cognitive Processes*, 18(4), 369-404.

[13] Tabak, W., Schreuder, R., and Baayen, R.H. (2005), Lexical statistics and lexical processing: semantic density, information complexity, sex and irregularity in Dutch, in M. Reis and S. Kepser (eds.), *Linguistic Evidence*, Berlin, Mouton de Gruyter, 529-555.

[14] Ford, M., Marslen-Wilson, W., and Davis, M. (2003), Morphology and frequency: contrasting methodologies, in H. Baayen and R. Schreuder (eds.), *Morphological Structure in Language Processing*, Berlin-New York, Mouton de Gruyter.

[15] Post, B., Marslen-Wilson, W., Randall, B., and Tyler, L.K. (2008), The processing of English regular inflections: Phonological cues to morphological structure, *Cognition*, 109, 1-17.

[16] Matthews, P.H. (1991), *Morphology*, Cambridge, Cambridge University Press.

[17] Corbett, G., and Fraser, N. (1993), Network Morphology: a DATR account of Russian nominal inflection, *Journal of Linguistics*, 29, 113-142.

[18] Pirrelli, V. (2000), *Paradigmi in Morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*, Pisa, Istituti Editoriali e Poligrafici Internazionali.

[19] Burzio, L. (2004), Paradigmatic and syntagmatic relations in italian verbal inflection, in J. Auger, J.C. Clements and B. Vance (eds.), *Contemporary Approaches to Romance Linguistics*, Amsterdam, John Benjamins.

[20] Blevins, J.P. (2006), Word-based morphology, *Journal of Linguistics*, 42, 531-573.

[21] Booij, G. (2010), *Construction Morphology*, Oxford, Oxford University Press.

[22] Altmann, G.T.M., and Kamide, Y. (1999), Incremental interpretation at verbs: restricting the domain of subsequent reference, *Cognition*, 73, 247-264

[23] Federmeier, K.D. (2007), Thinking ahead: the role and roots of prediction in language comprehension, *Psychophysiology*, 44, 491-505.

[24] Pickering, M.J. and Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105-110.

[25] Pirrelli, V., Ferro, M., and Calderone, B. (in press), Learning paradigms in time and space. Computational evidence from Romance languages, in *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*, in M. Goldbach, M.O. Hinzelin, M. Maiden and J.C. Smith (eds.), Oxford, Oxford University Press.

[26] Ferro M., Pezzulo G., and Pirrelli V. (2010), Morphology, Memory and the Mental Lexicon, in *Lingue e Linguaggio*, V. Pirrelli (ed.) *Interdisciplinary aspects to understanding word processing and storage*, Bologna, il Mulino.