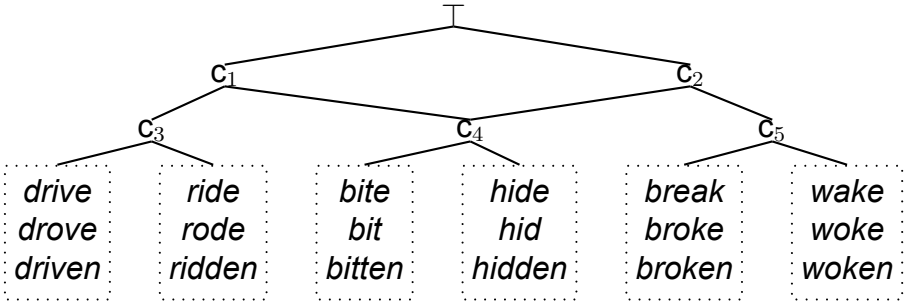
A comprehensive view on inflectional classification

Sacha Beniamine & Olivier Bonami
Université Paris Diderot, Laboratoire de linguistique formelle

Descriptions of inflection class systems take many forms. Pedagogical grammars are often content with giving a broad classification in major classes. At the other end of the spectrum, various studies (e.g. Stump and Finkel 2013) presuppose a classification into fine-grained micro-classes that exhaustively partition the set of lexemes. The two types of classifications can be linked by assuming a hierarchically-organized system of classes (Corbett and Fraser, 1993; Dressler and Thornton, 1996).

In recent years, various efforts have been made towards inferring automatically inflection class hierarchies from paradigms (Brown and Evans, 2012; Lee and Goldsmith, 2013; Bonami et al., 2014). While they use very different methodologies, all these approaches rely on the assumption that the hierarchy takes the shape of a tree. However, this is not the only option available. An alternative is to assume that the class system takes the form of a semi-lattice, where one subclass may belong to more than one superclass, as the figure below illustrates with sample English verbs.



We argue that semi-lattices are more appropriate to the modelling of inflection class systems, as they capture directly the phenomenon of heteroclisis (Stump, 2006), where different aspects of a lexeme's paradigm relate it to different classes. Class c_4 above is such a heteroclite class: it exhibits the same alternation between base form and participle as c_3 (Xaɪ$C \sim X$ɪCən), and the same alternation between past and participle as c_5 ($X \sim X$ən). We propose a computationally efficient method for inferring the semi-lattice of all groupings of lexemes sharing some inflectional characteristics. Technically, this is done by constructing a graph where each lexeme is linked to the alternation patterns (Bonami and Beniamine, 2015) it instantiates, and then introducing a class for each set of lexemes that is the scope of at least one pattern. The resulting semi-lattice can be interpreted as a monotonous inheritance hierarchy in the spirit of HPSG (Pollard and Sag, 1994), where each class indicates which patterns are simultaneously satisfied by all its members. However, the hierarchy is abstracted from paradigms of unsegmented words (Blevins, 2006) rather than constructed from other primitives, and represents exact and complete information on the system. Unlike previous proposals, it captures by design noncanonical paradigms involving defectivity or overabundance (Thornton, 2012).

In the presentation, we will show that the proposed method produces hierarchies that are a lot denser than hand-designed classifications by linguists, but a lot sparser than the theoretical maximum. For instance, on a dataset of 5000 fully conjugated French verbs (Bonami et al., 2014) exhibiting 73 distinct inflectional behaviours, we get 332 classes out of the theoretically conceivable $2^{73} \approx 10^{22}$.

References

Blevins, J. P. 2006. Word-based morphology. *Journal of Linguistics* 42:531–573.

Bonami, O. and Beniamine, S. 2015. Implicative structure and joint predictiveness. *In* V. Pirelli, C. Marzi, and M. Ferro (eds.), Word Structure and Word Usage. Proceedings of the NetWordS Final Conference.

Bonami, O., Caron, G., and Plancq, C. 2014. Construction d'un lexique flexionnel phonétisé libre du français. *In* F. Neveu, P. Blumenthal, L. Hriba, A. Gerstenberg, J. Meinschaefer, and S. Prévost (eds.), Actes du quatrième Congrès Mondial de Linguistique Française, pp. 2583–2596.

Brown, D. and Evans, R. 2012. Morphological complexity and unsupervised learning: validating Russian inflectional classes using high frequency data, pp. 135–162. *In* F. Kiefer, M. Ladányi, and P. Siptár (eds.), Current Issues in Morphological Theory: (Ir)regularity, analogy and frequency. John Benjamins, Amsterdam.

Corbett, G. G. and Fraser, N. M. 1993. Network morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics* 29:113–142.

Dressler, W. U. and Thornton, A. M. 1996. Italian nominal inflection. *Wiener Linguistische Gazette* 55-57:1–26.

Lee, J. and Goldsmith, J. A. 2013. Automatic morphological alignment and clustering. Presented at the 2nd American International Morphology Meeting.

Pollard, C. and Sag, I. A. 1994. Head-driven Phrase Structure Grammar. CSLI Publications; The University of Chicago Press, Stanford.

Stump, G. T. 2006. Heteroclisis and paradigm linkage. *Language* 82:279–322.

Stump, G. T. and Finkel, R. 2013. Morphological Typology: From Word to Paradigm. Cambridge University Press, Cambridge.

Thornton, A. M. 2012. Reduction and maintenance of overabundance. a case study on Italian verb paradigms. *Word Structure* 5:183–207.