Perspectives on morphological complexity

Anna Maria Di Sciullo
Université du Québec à Montréal

I discuss the notion of morphological complexity and compare the statistical approach to the deterministic approach. I focus on derivational morphology in Romance languages and in English. Notwithstanding the differences between these languages with respect to their overt morphological properties, I argue that they share substance-free morphological complexity, which can be reduced by factors external to the faculty of language (I-language). First I formulate the logical problem of (morphological) complexity, and show how statistical approaches account for the actual complexity of morphological utterances. Second, I then define the notion of substance-free (morphological) complexity and provide neurolinguistic evidence that supports its processing by the brain. Third, I discuss evidence from computational experiments that show that parsing complexity is reduced if recursive substance-free morphological complexity is linearized to the right. Finally, I draw some consequences for the import of statistical as opposed to I-language approaches to morphological complexity.

Morphological complexity can be evaluated on the basis of statistical calculi on the actual occurrences of affixes in a corpus (Bane 2007, Kolmogorov, 1965). Morphological complexity can also be evaluated on the basis of deeper properties of morphological structure, including hierarchical projections. Thus, in the statistical/probabilistic approach, one criteria of morphological complexity is the actual number of affixes available in a given language, and the number of possible combinations of these affixes with respect to roots/stems. A corpus-based analysis using Linguistica (Goldsmith 2001, 2006; Hu 2007) provides the ranking in (1) (from Bane 2007).

In the deterministic approach, complexity cannot be calculated via corpus-based analyses because such analyses do not necessarily rely on morphological sensorimotor (SM) substance. Interestingly, this kind of complexity is shared by languages which are apparently dissimilar with respect to the corpus-based statistical approach. Thus, according to (1), French has a lower percentage of complexity than Italian, and both languages have a higher percentage of complexity than English. To illustrate this point, I discuss the properties of verb-based derived categories in French, Italian and English in order to show that their derivation and interpretation require material that is not part of their overt morphological shape, (2)-(4).

I report the results of two psycholinguistic experiments on the acceptability of derived verbs and compounds in the languages under consideration. The results of these experiments show a significant difference in the priming/acceptability/complexity of morphological structure, whether the structures include material internal to the vP or not, (5).

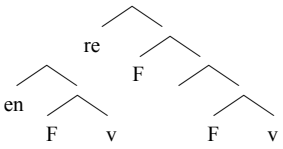
The complexity brought about by the absence of morphological material supporting conceptual information is also attested by computational experiments using a shift-reduce model for the derivation of forms such as *form-al-iz-able* (Di Sciullo and Fong 2005). The results show that derivational complexity grows exponentially if substance-free nodes precede the root, (6), (7). Assuming that the language design includes computational factors that reduce derivational complexity, including morphological complexity, these factors cannot be statistical or probabilistic in nature. It they were, it would not be possible to account for the fact that complexity brought about by SM substance-free projections can be reduced. Namely, substance-free complexity can be reduced if linearization proceeds by phases and is a function of the legibility of the edge of the phases.

(1) Rates of morphological complexity
(Bane 2007, 2008)

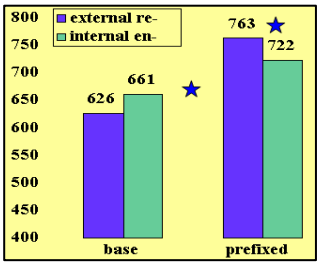
Latin 35.51%	English 16.88%
Hungarian 33.98%	Maori 13.62%
Italian 28.34%	Papiementu* 10.16%
Spanish 27.50%	Nigerian Pidgin* 9.80%
Icelandic 26.54%	Tok Pisin* 8.93 %
French 23.05%	Bislama* 5.38%
Danish 22.86%	Kituba* 3.40%
Swedish 21.85%	Solomon Pijin* 2.91%
German 20.40%	Haitian Creole* 2.58%
Dutch 19.58%	Vietnamese 0.05%

* = Creole/Pidgin

(2)



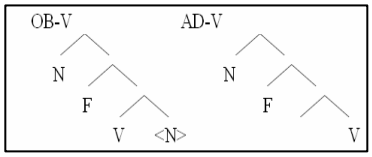
(3)



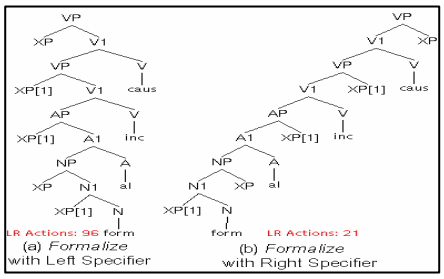
(4)

Mean RTs and SDs (prefixed and stem forms)		
	Mean RTs	Mean SDs
En stem	648	95
Re stem	628	80
En prefixed	724	97
Re prefixed	766	140

(5)



(6)



(7)

Word	Items	LR actions	
		Left Specifier	Right Specifier
form	1	8	6
read-able	2	19	11
simpl(e)-i(f)-y	3	47	16
form-al-i(z)-e	4	96	21
form-al-i(z)-(e) able	5	172	26

References

Bane, M. 2008. Quantifying and measuring morphological complexity. *Proceedings of the 26th West Coast Conference on Formal Linguistics*, pages 69–76.

Bane, M. 2007. Quantifying and Measuring Morphological *Complexity*. 26th *West Coast Conference on Formal Linguistics.*

Di Sciullo, A.M. and S. Fong 2005. Morpho-Syntax Parsing. In Anna Maria Di Sciullo (ed.) *UG and External Systems*. Amsterdam: John Benjamins. pp. 247-248.

Goldsmith, J. 2001. Unsupervised learning of the morphology of a natural language. Computational Linguistics 27.153–98.

Goldsmith, J. 2006. An algorithm for the unsupervised learning of morphology. *Natural Language Engineering* 12.1–19.

Kolmogorov, A. N. 1965. Three approaches to the quantitative definition of information. *Problems*
in Information Transmission 1.1–7.

Hu, Y, 2007. Topics in Unsupervised Language Learning. University of Chicago dissertation.