An Interdisciplinary Approach to Language Complexity: Mathematical and Computational Models

M. Dolores Jiménez López, Adrià Torrens Urrutia, Antoni Brosa Rodríguez & Susana Campillo Muñoz

Universitat Rovira i Virgili

Keywords: complexity, mathematical models, machine learning, Womb grammars, equicomplexity

This work focuses on the mathematical and computational study of linguistic complexity, a research area that has undergone a remarkable shift in recent decades. From an initial denial of the possibility of its measurement, supported by an equicomplexity perspective in the 20th century, there has been a growing interest in quantifying and comparing complexity among natural languages in the 21st century. In fact, the number of papers published in recent years on complexity both in the field of theoretical and applied linguistics highlights the interest in finding a method to calculate linguistic complexity and in trying to answer the question of whether all languages are equal in terms of complexity or, if on the contrary, they differ in their levels of complexity (Baechler and Seiler 2016; Baerman et al. 2015; Coloma 2017; Conti 2018; Di Domenico 2017 ; Kortman and Szmrecsanyi 2012; La Mantia et al. 2017; McWhorter 2012; Newmeyer and Preston 2014; Ortega and ZhaoHong 2017).

The examination of linguistic complexity is crucial in linguistics, shedding light on how languages vary in learning and processing. Computing has facilitated significant advances in this field, enabling large-scale analyses and model development. However, this transformation also poses theoretical and methodological questions that require careful consideration. Understanding both the advantages and disadvantages of computation in this context is crucial. This work aims to identify and critically evaluate key advantages of using computation in linguistic complexity analysis, explore associated disadvantages and limitations, including biases and data dependency, showcase instances illustrating how computation has enhanced our understanding of linguistic complexity, and provide concrete recommendations to address limitations and maximize the benefits of computation in this dynamically evolving field.

We will present different mathematical and computational models to calculate linguistic complexity based on machine learning tools and grammar frameworks such as Womb Grammars (Dahl and Miralles 2012). We will show how computation can enhance our understanding of linguistic complexity in diverse contexts. Moreover, we will provide insights into the advantages and challenges of computational approaches in linguistic complexity analysis, highlighting how computational models have opened new perspectives in language complexity computation. Our study aims to contribute to a better understanding of the capability of computation for large-scale analyses while fostering a clearer awareness of computational limitations. As a result, this work will contribute to highlighting the importance of linguistic complexity analysis and its relevance in current linguistic research, challenging the dogma of equicomplexity among languages by providing mathematical models that enable the calculation of the relative complexity of languages.

ACKNOWLEDGMENTS: Grant PID2020-120158GB-I00 funded by MICIU/AEI /10.13039/501100011033

REFERENCES

Baechler, R. and Seiler, G. (2016), *Complexity, Isolation, and Variation*. Berlin: De Gruyter.

Baerman, B., Brown, D. y Corbett, G.G. (2015), Understanding and measuring morphological complexity. Oxford: Oxford University Press.

Coloma, G. (2017), *La Complejidad de los Idiomas*. Berlin: Peter Lang.

Conti Jiménez, C. (2018), Complejidad lingüística: Orígenes y revisión crítica del concepto de lengua compleja. Berna: Peter Lang

Dahl, V. and Miralles, E. (2012), Womb grammars: constraint solving for grammar induction, in J. Sneyers y T. Frühwirth (eds.), *Proceedings of the 9th Workshop on Constraint Handling Rules*, Technical Report CW 624. Roskilde University, 32-40.

Di Domenico, E. (2017), *Syntactic Complexity from a Language Acquisition Perspective.* Newcastle upon Tyne: Cambridge Scholars Publishing.

Kortmann, B. and Szmrecsanyi, B. (2012), *Linguistic Complexity: Second Language Acquisiton, Indigenization, Contact*. Berlin: Mouton de Gruyter.

La Mantia, F., Licata, I., Perconti, P. (eds.) (2017), *Language in Complexity. The Emerging Meaning*. Berlin: Springer.

McWhorter, J. (2001), The world's simplest grammars are creole grammars, *Linguistic Typology 6, 125–166.*

McWhorter, J. (2012), *Linguistic Simplicity and Complexity: Why do Languages Undress?* Berlin: Mouton de Gruyter.

Newmeyer, F.J. and Preston, L.B. (eds.) (2014), *Measuring Grammatical Complexity*. Oxford: Oxford Univesity Press.

Ortega, L. and Han ZhaoHong (2017), *Complexity Theory and Language Development* Amsterdam: John Benjamins.