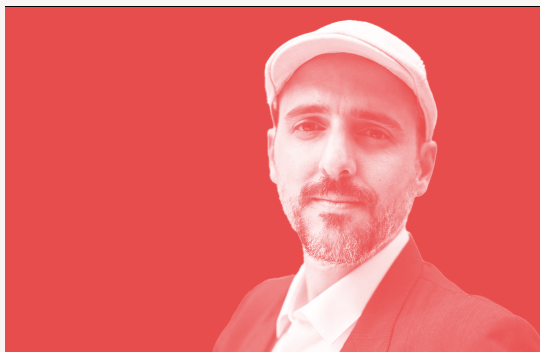


Word boundaries and the morphology-syntax trade-off

Pablo Mosteiro

Universiteit Utrecht



This work investigates the relationship between syntax and morphology in natural languages, focusing on the relation between the amount of information stored by word structure on the one hand, and word order on the other. In previous work, a trade-off between these was observed in a large corpus covering over a thousand languages, suggesting a dynamic “division of labor” between syntax and morphology, as well as yielding proof for the efficient coding of information in language. In contrast, we find that the trade-off can be explained by differing conventions in orthographic word boundaries. We do so by redefining word boundaries within languages either by increasing or decreasing the domain of wordhood implied by orthographic words. Namely, we paste frequent word-pairs together and split words into their frequently occurring component parts. These interventions yield the same trade-off within languages across word domains as what is observed across languages in the orthographic word domain. This allows us to conclude that the original claims on syntax-morphology trade-offs were spurious and that, more importantly, there does not seem to exist a privileged wordhood domain where within- and across-word regularities yield an optimal or optimized amount of information.

**Dijous, 17 d'octubre
de 2024 - 16h**

Aula 0.2

Facultat de Filologia i Comunicació