# Syntactic change and DLM in German: a corpus study

This study argues that the variation in the placement of relative clauses (RC) in German can be explained by the principle of dependency length minimization (DLM), which states that languages tend to place syntactically and semantically related words close together (e.g., Gibson, 2000, Futrell et al., 2015). Although Gildea & Temperley (2010) ascribe only a weak effect of DLM to German grammar, my diachronic corpus study will show that we can see a strong effect of DLM by looking at the RC position in the history of written German. I will testify to a two-stage change: The syntactic complexity reflected by a high number of intraposed RC first increases in the 17th and 18th centuries due to the growing standardization of the written language, which favors verb-final structures and reaches its peak at the end of the 18th century (Admoni, 1967), before it decreases continuously so that extraposition becomes the most frequent word order.

According to Hawkins (1994, 2004), word order variation in sentence production can be explained by processing economy constraints. Specifically, he states that, given alternative word orders within a construction, the one that allows faster recognition of the immediate constituents is expected to be preferred. Gibson (1998, 2000) takes a similar approach concerning sentence comprehension. He establishes that the memory cost is higher the more incomplete syntactic dependencies one has to keep track of during sentence processing. In German word order, auxiliary and main verb are placed discontinuously in a sentence. According to DLM, there should be a tendency to reduce the distance between the verbs by outsourcing heavy NPs behind the finite verb to the end of the sentence to avoid an overload of the working memory capacity.

Addressing the question of whether DLM plays a role in explaining the changing RC position in German, I investigate the placement of heavy NPs and PPs with relative clause modifiers from newspaper texts from 1600 until Present-day German. The relevant structures are intraposition where the RC is placed adjacent to its head noun (1) and extraposition where the nominal head appears preverbally and the RC occurs at the right edge of the sentence (2).

(1)  Von deß Orators Leuten **sollen** 2. [$_{PP}$ vnter einem tumult/ [so die Türcken deß Nachts in jhrem Losament angefangen]] **vmbkommen sein** (1609: Relation)
   'Two people of the orator are said to have been killed during a tumult that the Turks started in their accommodation at night.'

(2)  sonsten **weren** in Spannia auch [$_{NP}$ Ampassatores vom König Matthiasen vnd Hertzogen von Savoia] **angelangt** / [so beim König schon Audienz gehabt] (1609: Relation)
   'Furthermore, ambassadors of King Matthias and of the Duke of Savoy had also arrived in Spain, who had already had an audience with the king.'

An intraposed word order with a long dependency (i.e., a high amount of language material between the verbs) leads to high working memory costs and therefore runs into the risk of processing difficulties. Differently, extraposition can facilitate sentence processing because the dependency length (DL) between the verbs is minimized. Complicating matters, however, is that the DL between RC and its antecedent is increased with extraposition, thus, presenting potentially competing motivations.

To measure syntactic complexity, I consider the factors (i) RC length, (ii) distance between the discontinuous verbs, and (iii) distance between the RC and its antecedent. My results show a strong effect of DLM on the development of German word order preferences: both DLs (between RC and its antecedent and between the separated verbs) decrease over time significantly. On the one hand, extraposition becomes the most frequent word order which leads to reduced distances between the discontinuous verbs over time. On the other hand, the DL between the extraposed RC and the antecedent decreases because it becomes the norm that no more than two words can occur here. This can be explained by the processing pressure in spoken modalities that influence the written standard in newspapers over time increasingly.

# References

Admoni, Wladimir G. 1967. Der Umfang und die Gestaltungsmittel des Satzes in der deutschen Literatursprache bis zum Ende des 18. Jahrhunderts. *Beiträge zur Geschichte der deutschen Sprache und Literatur, 89*, 144–199.

Futrell, Richard, Kyle Mahowald, & Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336-10341.

Gibson, Edward. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition, 68*(1), 1–76.

Gibson, Edward. 2000. The dependency locality theory: A distance-based theory of linguistic complexity. *Image, language, brain*, 95–126.

Gildea, Daniel, and David Temperley. 2010. Do grammars minimize dependency length? *Cognitive Science 34*, 286–310.

Hawkins, John A. 1994. A Performance Theory of Order and Constituency. *Journal of Linguistics*, *33*(1), 171–212.

Hawkins, John A. 2004. Efficiency and Complexity in Grammars. Oxford: Oxford University Press.