The Effect of Noun Phrase Complexity in Scientific Texts on Reading Times of Experts and Novices

Isabell Landwehr (Saarland University), Marie-Pauline Krielke (Saarland University), Stefania Degaetano-Ortlieb (Saarland University)
isabell.landwehr@uni-saarland.de

We investigate how different types of noun phrase (NP) complexity in scientific texts affect the reading times of experts and novices, for both in-domain and out-of-domain texts. The use of complex NPs is a key feature of scientific writing (Biber & Gray, 2011). For sentence processing, NP complexity can pose various challenges: More complex structures often include longer dependencies between head and dependent, increasing the integration cost of syntactic elements (cf. Dependency Locality Theory, Gibson, 1998). Moreover, complex NPs allow for information to be transmitted in a more compressed way increasing implicitness (Biber & Gray, 2010): Logical relations between the constituents of a compound remain implicit. Previous eye-tracking experiments show that increased complexity correlates with increased reading times (e.g. Just & Carpenter, 1980, for scientific texts). Individual reader characteristics, e.g. background knowledge and experience, also influence reading comprehension (Kendeou & Van den Broek, 2007). This is particularly the case for scientific texts, typically targeted at an expert audience (Halliday, 1988). Previous studies have considered word frequency or novelty (Just & Carpenter, 1980), dependency locality (Demberg & Keller, 2008) or terminology (Škrjanec et al., 2023) as complexity features.

We consider grammatical complexity by looking at structural compression (Biber & Gray 2016, p. 207). In particular, we analyze (a) different types of NP modification, i.e. different degrees of compression (see Table 1), and (b) differing internal structure (see Table 2). We use PoTeC (Jakobi et al., 2024), a German naturalistic eye-tracking-while-reading corpus of university students (novices: BA, experts: MA, PhD) of biology or physics reading in/out-of-domain textbooks. We foresee an effect of expertise, given higher domain knowledge for experts vs. novices: increased processing difficulty for novices for NPs with higher degree of compression and more complex internal structure, such as compounds, compared to e.g. nouns modified by a genitive construction. Additionally, experts are likely to outperform novices when reading texts from other scientific fields as their general scientific reading competence provides an extra advantage.

We fit linear mixed effects regression models using the *lme4* (Bates et al., 2015) package in R (R Core Team, 2023). Our dependent variables are first-pass reading time, total fixation time and total no. of incoming regressions. Our predictors are NP modification type or internal structure, and reader expertise, allowing us to model the effect of NP complexity considering reader's level of expertise and domain familiarity. As in previous work, we control for word length, type frequency, technicality of a term and surprisal. We also include an interaction of complexity and expertise, technicality and expertise as well as by-subject and by-word random effects. As a result, we aim to highlight the role of NP complexity on processing difficulty, and its interaction with readers' domain expertise.

Table 1: Compounds with different types of modification

Modification type (degree of compression)	Example
Nominal compound (higher compression) vs. modification by adjective (lower compression)	Wildtypprotein vs. endogenes zelluläres Protein
Nominal compound (higher compression) vs. modification by genitive construction (lower compression)	Phosphatverarmungszonen vs. Wellenvektor des Elektrons

Table 2: Compounds with differing internal structure

Internal structure	Example
Compound with one vs. two dependents	Energieminimum vs. Phosphatverarmungszonen

References

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Biber, D. & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes 9*, 2-20. https://doi.org/10.1016/j.jeap.2010.01.001

Biber, D. & Gray, B. (2011). Grammatical change in the noun phrase: The influence of written language use. *English Language and Linguistics 15* (2), 223-250. https://doi.org/10.1017/S1360674311000025

Biber, D. & Gray, B. (2016). *Grammatical Complexity in Academic English.* Cambridge University Press. https://doi.org/10.1017/CBO9780511920776

Demberg, V. & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition 109*, 193-210. https://doi.org/10.1016/j.cognition.2008.07.008

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

Halliday, M. A. K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of Written English: Situational Factors and Linguistic Features* (pp. 162-178). Pinter Publishers.

Jakobi, D. N., Kern, T., Reich, D. R., Haller, P. & Jäger, L. A. (2024). *PoTeC: A German naturalistic eye-tracking-while-reading corpus*. https://doi.org/10.17605/OSF.IO/DN5HP

Just, M. A. & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review 87*(4), 329-354. https://doi.org/10.1037/0033-295X.87.4.329

Kendeou, P. & Van Den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory & Cognition 35*(7), 1567-1577. https://doi.org/10.3758/BF03193491

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Škrjanec, I., Broy, F. Y. & Demberg, V. (2023). Expert-adapted language models improve the fit to reading times. *Procedia Computer Science 225*, 3488-3497. https://doi.org/10.1016/j.procs.2023.10.344