Noun+noun compounds and verbal complements as non-normalised features in Late Modern English scientific translations

1 author:

Katrin Menzel
University of Mannheim
31 PUBLICATIONS **117** CITATIONS

Noun+noun compounds and verbal complements as non-normalised features in Late Modern English scientific translations

Katrin Menzel, University of Mannheim, Department of English, katrin.menzel@uni-mannheim.de

This paper presents a study on the usage of noun+noun compounds and verbal complement structures in 18th century scientific articles in the Royal Society Corpus (RSC) comparing translated to non-translated English texts. Departing from the hypothesis that the translations will conform stronger to traditional patterns of the English language, the analysis shows that these historical translations and non-translated texts are similarly marked by the ongoing reorganisation of the noun phrase, but translations contain more innovative complementation patterns. Additionally, a surprisal analysis shows that the analysed patterns tend to occur in more predictable and conventionalised contexts in non-translated texts than in translations.

1. Introduction

This paper presents a study on the usage of noun+noun compounds and verbal complement structures in Late Modern English (LModE)

scientific articles in the Royal Society Corpus (RSC, V6.0 / 7.0, Fischer et al. 2020, time span covered by the entire corpus: 1665-1996, ca. 48,000 texts from journals such as the *Philosophical Transactions* and *Proceedings* of the Royal Society of London).[1] The analysis focusses particularly on comparing 18th century English translations of scientific articles to non-translated English articles in the dataset. It therefore addresses a general research gap with regard to corpus-based diachronic research on translations and with regard to English as a target language as it has received considerably less attention in translation studies than English as a source language.

Noun+noun open compounds used as terms (e.g. *antagonist muscles, burning lens*) and *-ing*-complements after verbs taking over functions of *to*-infinitives in LModE (e.g. *omit telling* vs. *omit to tell*) – which, in turn, had become more typical alternatives of complement *that*-clauses from Middle English onwards according to the 'Great Complement Shift' hypothesis – are innovative patterns that gained in frequency during LModE (Biber & Gray 2016, Rohdenburg 2006: 144). This paper departs from the normalisation hypothesis (Baker 1996), a T-universal in the sense of Chesterman's (2004) use of the term, with the expectation that translated 18th century research articles will be characterised by more 'conservative' and conventional patterns than non-translated articles. If the normalisation hypothesis is universally valid, the 18th century translations in the RSC will conform stronger to traditional grammatical patterns of the English language up to the point of exaggerating them compared to other scientific texts from the same time span. The translations should contain fewer noun+noun compounds, fewer *-ing*-complements after verbs, and more traditional structures.

Besides comparing frequency distributions over time, surprisal-based measures from an information-theoretic framework are compared (cf. Hale 2001, Levy 2008). The RSC is annotated for surprisal values (S), calculated as the negative log (base 2) probability (p) of each token (t) in the RSC given its preceding context of three tokens measured in bits of information: $S(t_i) = -\log_2 p(t_i|(t_{i-1}\ t_{i-2}\ t_{i-3}))$. These values are indicators of the amount of information and contextual predictability

---

of the components of the analysed patterns.

The analysis shows that the translations are less 'normalised' than expected. They resemble non-translated texts with regard to the analysed features or use more innovative and fewer traditional patterns.

2. Identifying translations in the RSC

The article titles are a valuable source for identifying translated English articles in the RSC. Additionally, the metadata in the forthcoming RSC version 7.0 contain information on the 'author roles' for most texts (e.g. author, reviewer, translator). When we combine the results from queries for translation-related information in the text titles or in the metadata and delete the duplicates, we obtain a list of 198 translated texts from 1668-1991 (Figure 1). Most translated articles in the corpus were published during the 18th century, and most were translations from French (Figure 2).
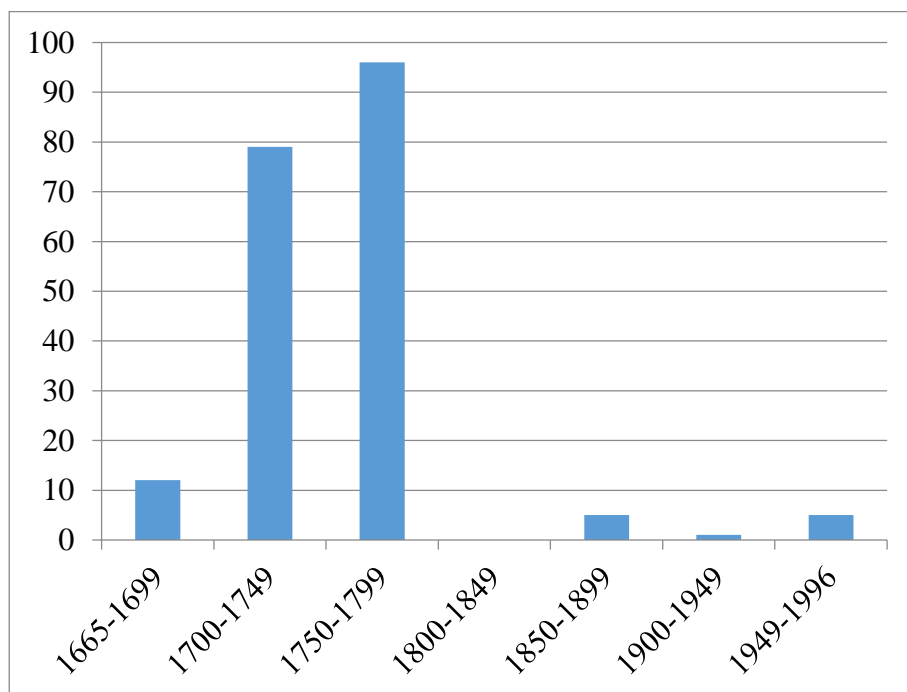

Figure 1: Translated articles per 50-year periods (RSC 7.0) [2]

---

[2] Generally, the source texts are not included in the journal articles for a direct comparison with their translations, apart from a few articles that contain the original text and its English translation in one document. These texts with both
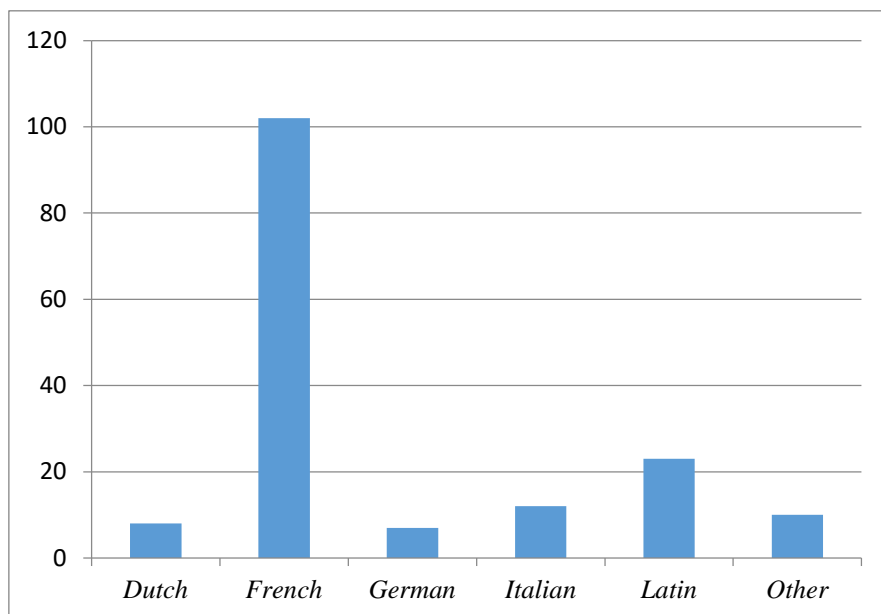
Figure 2: Number of translations with information on their source language

These translations were typically produced by scientists from the Royal Society's network who did not have translation tasks as their main occupation. One might think that translations played only a marginal role in the overall RSC. Nevertheless, these texts represent certain fields in the respective time spans with important research centres abroad (e.g. astronomy, biology, chemistry in the 17th and 18th centuries).

3. Analysis of translated and non-translated 18th century research articles

As most translations in the RSC were produced during the 18th century and few were produced during the 19th century, the analysis

the original and the translation will be excluded in the analysis in Section 3 as they have been deleted from most RSC versions to enhance the annotations. However they may be used in a follow-up qualitative analysis focussing on possible source language interferences.

concentrates on this first period of LModE. The 18th century translations are treated as a specific subcorpus (463,153 tokens, 168 texts) that is contrasted with another subcorpus of randomly selected non-translated articles from the same time span (463,210 tokens, 216 texts).[3]

The 18th century translations and the subcorpus with non-translated 18th century texts are queried with CQPweb (Hardie, 2012) for frequency information on noun+noun sequences used as open compounds and *-ing*-complements vs. *to*-infinitives after verbs that occur with both forms in LModE (cf. examples above).[4] Then the surprisal values of the components of the compounds and of the verb forms in the complements are extracted and compared. The results show that translations do not lag behind with regard the use of noun+noun compounds or *-ing*-complements after verbs compared to other 18th century RSC texts.

Figure 3 shows the frequencies of noun+noun sequences in the subcorpora. They occur slightly more often in translations.

---

[3] One might consider more specific variables with regard to the texts such as their topics, length or source language (even if the source texts are not included in the corpus), or compare specific time-spans within the 18th century. However, a larger size of the dataset would probably be necessary for looking into these further details with the aim of going beyond example-based arguments.

[4] While the queries for verbal complement structures only needed some refinements to exclude a few irrelevant patterns, queries for noun+noun sequences needed many refinements to exclude a high number and variety of patterns that do not function as compounds or were wrongly tagged as nouns in the 18th century data. Noun sequences including proper nouns were not included in the analysis.
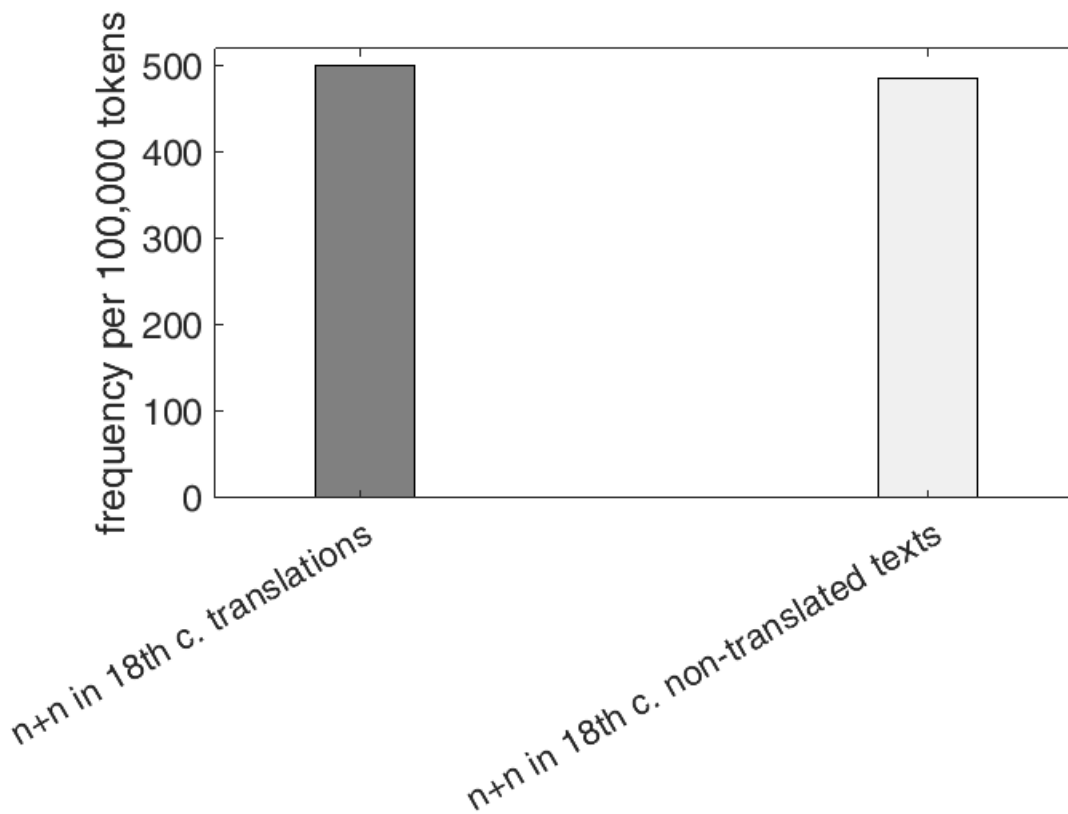
Figure 3: Noun+noun compounds in translated vs. non-translated texts

Figure 4 compares the range of surprisal values of nouns in this pattern in the two subcorpora (summarising the values for modifying and head nouns). The boxes look similar, but the median in the translations is 14.5 compared to 13.2 in the non-translated texts – a small, but statistically significant difference, which we can conclude from the plotted notches that represent the confidence interval around the median.[5] This indicates that, on average, this pattern is used in less conventionalised and less predictable contexts in 18th century translations than in non-translated texts.

---

[5] Notches display the variability of the median between samples. The width of a notch is computed by MATLAB so that boxes whose notches do not overlap have different medians at the 5% significance level. Comparing box plot medians is analogous to the *t* test used for means. (https://de.mathworks.com/help/stats/boxplot.html Accessed 22/04/2024)
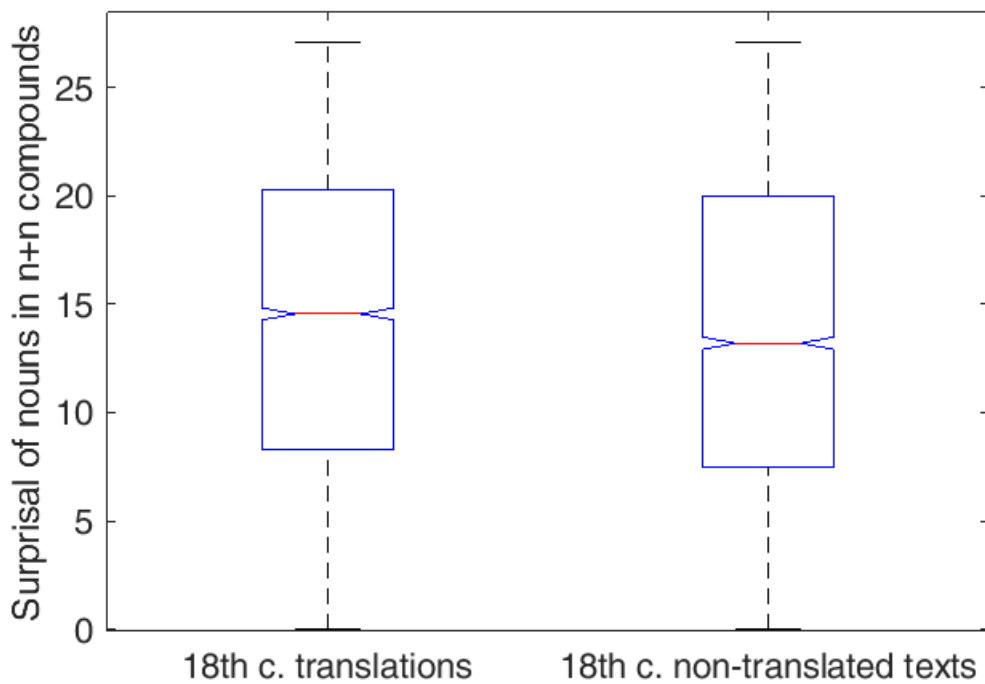
Figure 4: Summarised ranges of surprisal of noun+noun compounds

Figure 5 compares *-ing*-complements and *to*-infinitive structures after verbs that occur with both (*-ing*-complement or *to*-infinitive) in the texts from the two subcorpora (e.g. *attempt, continue, omit, pretend, propose, refuse, seem*). Verbal complements with *to*-infinitives are generally more frequent in this time span. However, the more traditional *to*-infinitive complement is more frequent in non-translated texts compared to translations, while *-ing*-complements after verbs as a more innovative pattern in LModE are more frequent in translations.
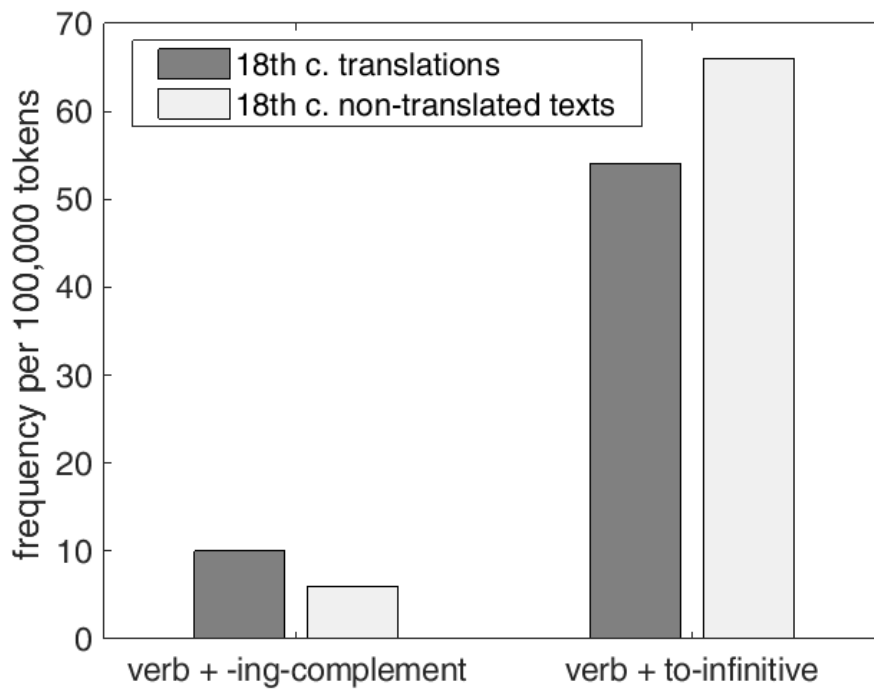
Figure 5: Frequencies of verb + -ing and verb + to-infinitive complements

The surprisal analysis of the complementation patterns does not lead to a clear-cut conclusion. Figure 6 shows the surprisal range of *-ing*-forms and infinitives in verbal complements. The surprisal of *-ing*-forms is slightly higher in non-translated texts, but the medians do not differ in a statistically significant way as the 18th century RSC texts generally do not contain numerous instances of this pattern yet (cf. Figure 5).
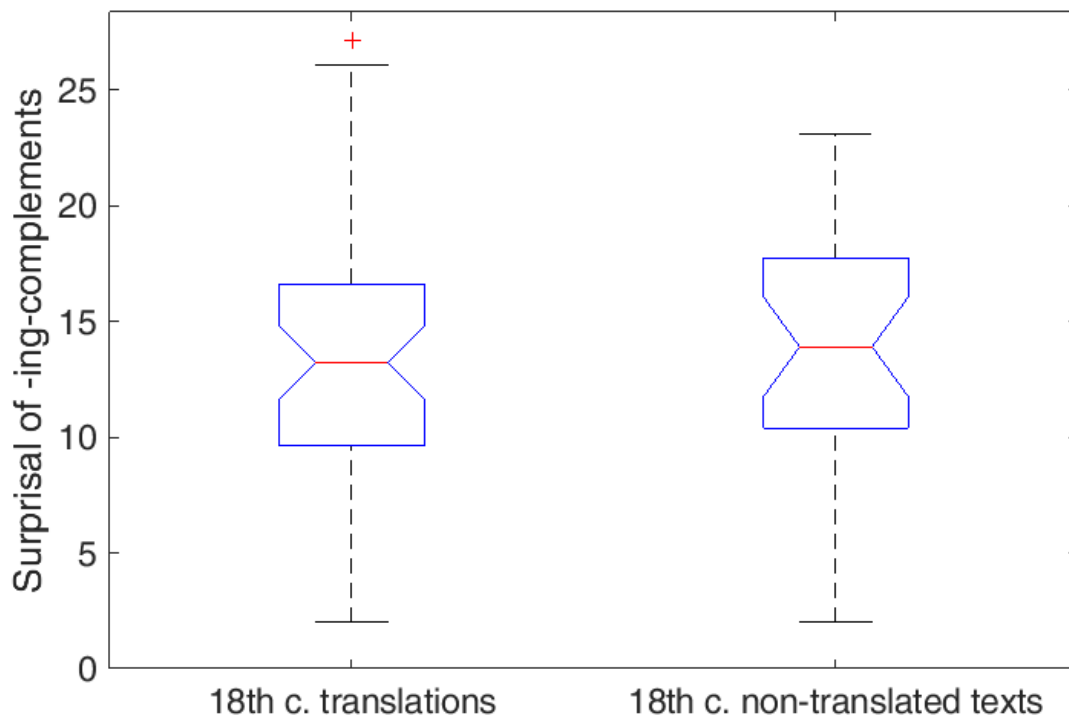
Figure 6: Surprisal of -ing-forms used as complements after verbs

The boxes in Figure 7 show that the surprisal of infinitives used as verbal complements is generally higher in translations than in non-translated texts and that the medians differ in a statistically significant way (6.5 translations, 4.3 in non-translated texts).
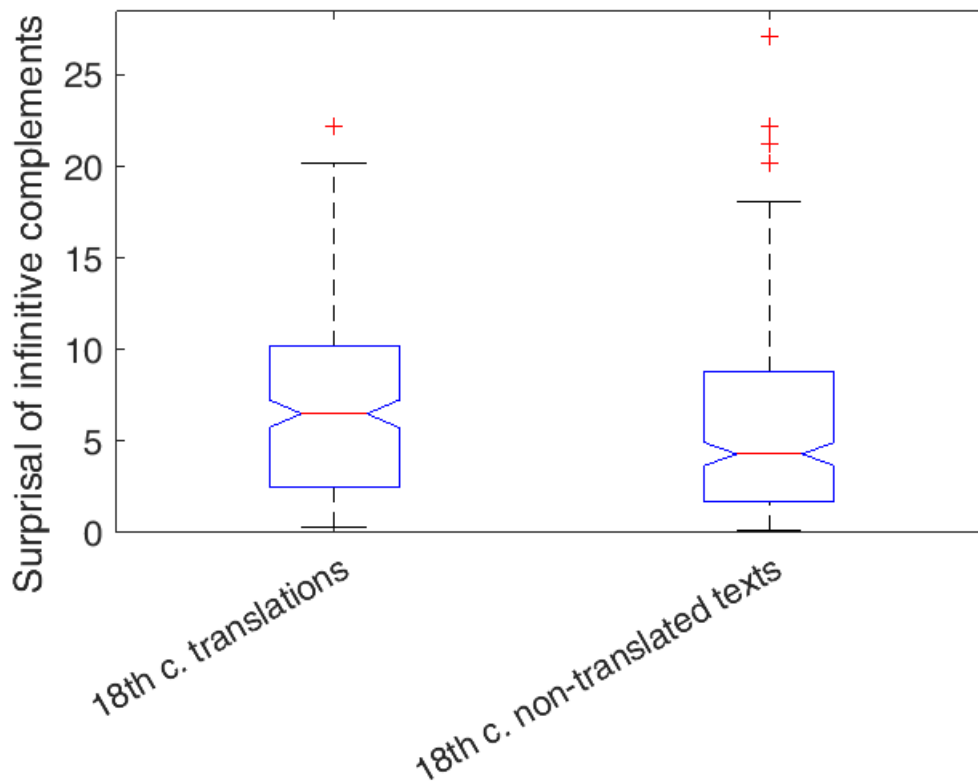
Figure 7: Surprisal of to-infinitive verb forms used as complements after verbs

This indicates that this pattern, which occurs less often in translations, is used in these texts after less predictable and less conventionalised preceding contexts. It is more frequent in non-translated texts and occurs in more predictable and conventionalised contexts there signalling lower information density and cognitive load of this pattern compared to translations.

4. Conclusions

Overall, the analysis shows that normalisation or conservatism is not a prevalent translation norm in 18th century translated scientific texts. The translations and non-translated texts are similarly marked by the ongoing reorganisation of the noun phrase, but translations show more innovative complementation patterns. Moreover, the analysed patterns tend to occur in more predictable and conventionalised contexts in the non-translated texts.

References

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In: H. Somers (ed.), *Terminology, LSP and translation. Studies in language engineering in honour of Juan C. Sager*. 175–186. Amsterdam: Benjamins.

Biber, D., & Gray, B. (2016). *Grammatical complexity in Academic English – Linguistic change in writing.* Cambridge: Cambridge University Press.

Chesterman, A. (2004). Beyond the particular. In A. Mauranen and P. Kuyamaki (eds.), *Translation universals: Do they exist?* Amsterdam: John Benjamins, 33–49.

Fischer, S., Knappen, J., Menzel, K. & Teich, E. (2020). The Royal Society Corpus 6.0. Providing 300+ years of scientific writing for humanistic study. *Proceedings of LREC 2020,* Marseille, May 2020, 794–802.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the 2nd meeting of the North American chapter of the association for computational linguistics on language technologies*, Pittsburgh, June 2001, 1–8.

Hardie, A. (2012). CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics,* 17: 380–409.

Levy, R.P. (2008). Expectation-based syntactic comprehension. *Cognition*, 106 (3): 1126–1177.

Rohdenburg, G. (2006). The role of functional constraints in the evolution of the English complementation system. In: C. Dalton-Puffer, D. Kastovsky, N. Ritt and H. Schendl (eds.), *Syntax, style and grammatical norms: English from 1500–2000*, 143–166. Bern: Peter Lang.