A Rational Speech Act model of cross-linguistic differences in pronoun resolution preferences

Miriam Schulz, Heather Burnett, & Barbara Hemforth (CNRS-Univ. Paris Diderot)
barbara.hemforth@linguist.univ-paris-diderot.fr

An important current research question in psycholinguistics concerns the mechanisms through which different interpretations of superficially similar constructions can arise across languages. [1] observed cross-linguistic variation in sentences like (1). In English, comprehenders prefer to resolve the pronoun to the subject, while, in French, they are most likely to interpret the pronoun as referring to the object.

(1) a. The postman called the streetsweeper before he went home.
   b. Le facteur a appelé le balayeur avant qu'il rentre à la maison.

We propose a computational model of differences in pronominal resolution preferences between English and French active sentences within the RSA framework that takes into account differences in the inventory of syntactic constructions between them. The model predicts the patterns of pronominal reference observed in experiments based on a high prior probability of next mention of subjects (see [2]) and the availability of an alternative non-finite construction in the two languages (2, 3) that unambiguously takes subject antecedents.

(2) The postman called the streetsweeper a. before/b. after going home.
(3) a. Le facteur a appelé le balayeur avant de rentrer à la maison.
   b. Le facteur a appelé le balayeur après être rentré à la maison.

A corpus study of English (COCA) and French (Frantext, Est Républicain, ESLO) shows that alternative non-finite constructions for sentences like (1) are much less frequent in spoken English than French. The corpus study further extends the picture by distinguishing between spoken, literary and journalistic language as well as including *after* (2b, 3b) as a second connector besides *before* which has been studied in previous research. Logistic regressions (glm) show an independent significant impact of language, genre and connector as well as interactions between connector and genre on the frequency of occurrence of the finite with respect to the non-finite construction (Figures1a,b, ps < .001).

A Rational Speech Act model [3] of the interpretation of pronouns in ambiguous utterances, integrating an assumed general preference for subject antecedents across languages (proportion of .80 subject antecedents) as well as spoken corpus frequencies as language-specific costs on the use of each construction makes empirical predictions for pronoun resolution preferences in French and English for finite subclauses with *avant* or *before* as well as with *after* or *après*. Language-specific cost is calculated as the natural log of the proportions of the finite and non-finite constructions in the spoken English and French corpora.

The model predicts an object preference for French sentences with *avant* as well as with *après* (Figure 2a,b) and a subject preference for English across connectors, though more so for *before* given the higher frequency of the non-finite alternative for *after* (see [4] for details).

To test model predictions, we carried out a crosslinguistic experiment for 16 sentences like (1a,b) using a cloze task where participants had to fill in a gap after each sentence as in "_____ went home." 100 French and 60 English participants provided antecedent choices with conditions distributed across lists following a Latin Square design (Fig.3). Logistic regression analyses show significant effects of Language (with fewer N2=object choices in English) and an interaction of Language and Connector with more N2 choices for *after* than *before* in English but fewer for *après* in French (all ps<.001). While the higher number of N2 choices for *after* in English is predicted by the higher relative frequency of non-finite constructions with *after*, the lower number of N2 choices in French was not predicted (see Fig. 2). We assume that the generally low proportion of *après* constructions in French but not of *after* in English plays a role here (Fig. 4). Only if both alternatives are sufficiently frequent in a language, the relative cost will have an impact, otherwise the general subject preference takes over.
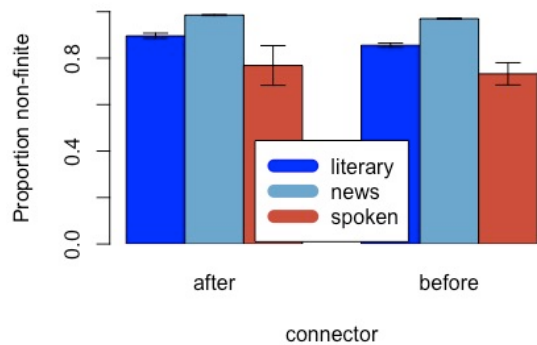
Figure 1a: Proportion of non-finite constructions for *avant* and *après* in French
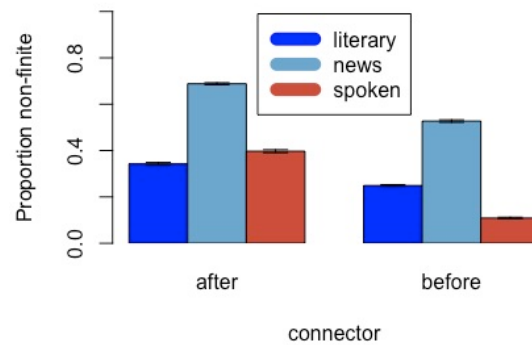
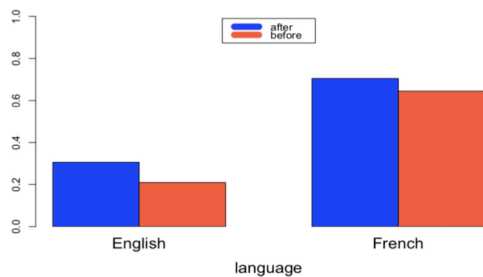Figure 1b: Proportion of non-finite constructions for *before* and *after* in English

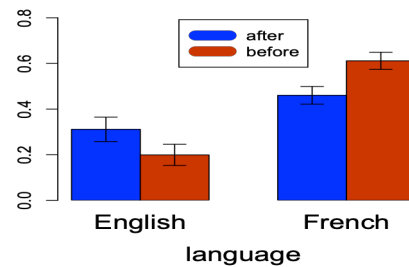Figure 2: Model predictions for proportion of objects choices

Figure 3: Antecedent choices in English and French (proportion object choices) N2=object)
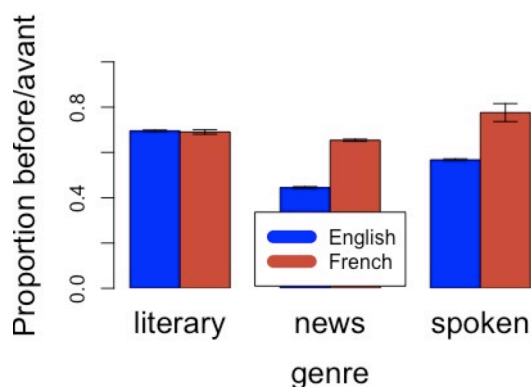

Figure 4: Proportion of *before/avant* out of all before/avant and after/après combined

[1] Hemforth, et al. (2010). Language specific preferences in anaphor resolution: Exposure or gricean maxims? In S. Ohlsson, S & R. Catrambone (Eds.), Proceedings of the 32nd Annual Conference of the Cognitive Science Society, 2218–2223, Portland, OR.
[2] Arnold , J. (2001) The Effect of Thematic Roles on Pronoun Use and Frequency of Reference Continuation, Discourse Processes, 31:2, 137-162
[3] Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science* 336, 998.
[4] Schulz et al.(submitted). Corpus, Experimental and Modelling Investigations of Cross-linguistic Differences in Pronoun Resolution Preferences.