

A Bayesian Game-Theoretic Approach to Cross-Linguistic Variation

Heather Burnett & Barbara Hemforth
(LLF, CNRS-Université Paris Diderot)

1. Introduction. An important current research question in psycholinguistics concerns the mechanisms through which different interpretations of superficially similar constructions can arise across languages (see Frazier & Clifton 1995, Grillo & Costa 2015 for cross-linguistic variation in the interpretation of relative clauses; Carminati 2002, Hemforth et al. 2010, de la Fuente et al. 2016 for variation in pronoun resolution). Many of these researchers suggest that Gricean-type principles may be underlying the observed differences; however, they do not provide formal accounts of the relationship between interactive reasoning, syntactic variation and pragmatic interpretation. This paper argues that Bayesian signaling game models, particularly *Iterated Best Response (IBR)* or *Rational Speech Act (RSA)* models (Franke 2009, Frank & Goodman 2012 et seq.), can significantly contribute to making these proposals maximally explicit. Although these models have been shown to be useful in the analysis of many pragmatic phenomena in English, the potential of this framework for analyzing cross-linguistic pragmatic variation has yet to be explored. In IBR/RSA models, speakers' actions (and consequently listeners' interpretations) are optimized according to the inventory of syntactic forms available in the language. Therefore, this architecture is ideal for capturing the link between syntactic variation and variation in interpretation.

We build a simple RSA model of differences in pronominal resolution preferences between German, English and French that takes into account differences in the inventory of syntactic constructions between them, and we show how our model straightforwardly predicts the patterns of pronominal reference observed in psycholinguistic experiments. We therefore conclude that game-theoretic models constitute valuable tools for investigating the link between the syntactic properties of a language and the pragmatic reasoning processes of its speakers.

2. Cross-linguistic data. Consider an utterance with two possible referents (subject and object) (1). Using both visual world and questionnaire methodology, Author et al. 2010 and Baumann et al. 2014 have show that, when such an utterance is followed by a separate sentence containing a pronoun (2), German, French and English listeners highly prefer to interpret the pronoun as referring to the subject rather than the object.

- | | | | |
|-----|----|---|----------------|
| (1) | a. | Der Briefträger hat den Straßenfeger getroffen... | German |
| | b. | The postman met the street sweeper ... | English |
| | c. | Le facteur a rencontré le balayeur ... | French |
| (2) | a. | . Dann ging er nach Hause. | German |
| | b. | . Then he went home. | English |
| | c. | . Puis il est rentré à la maison. | French |

However, these authors also observe cross-linguistic variation when the following utterance is an adjunct on the main clause (3). In German and English, listeners still prefer to resolve the pronoun to the subject (see also Bouma & Hopp 2007, Kehler & Rohde 2016). However, in French, they are most likely to interpret the pronoun as referring to the object (see also Colonna et al. 2012).

- | | | | |
|-----|----|---|----------------|
| (3) | a. | ... bevor er nach Hause ging. | German |
| | b. | ... before he went home. | English |
| | c. | ... avant qu' il rentre à la maison. | French |

For example, in Hemforth et al. 2010's questionnaire, the percentage of subject interpre-

tations across and within sentences is shown in the table below:

LANGUAGE	BETWEEN SENTENCE	WITHIN SENTENCE
German	95%	85%
English	95%	70%
French	95%	20%

3. An RSA model for German/English/French. RSA models formalize aspects of Gricean reasoning in terms of Bayesian signaling games. In a signaling game, there are two players: speaker (S) and listener (L). S has a fact about the world that they want to communicate to L. To model the pronoun resolution data, we will assume that the space of propositions under consideration consists of *The individual denoted by the subject of (1) went home* (p_S) and *The individual denoted by the object of (1) went home*. (p_O). S has a set of interpreted syntactic forms that they can choose from to send to L. As Hemforth et al. 2010 observe, German, English and French crucially differ both on the inventory of forms in the language and on the patterns of use of syntactic variants. Unlike German, both English and French possess variants of (3) containing a null PRO (4), which is obligatorily interpreted as referring *de se* to the subject (Chierchia 1989).

(4) ... **before going** home. (**Eng.**) ... **avant de rentrer** à la maison. (**Fren.**)

Furthermore, English and French differ in the relative frequency of the PRO form: in corpus studies of English, the overt pronoun form was found to be 4.32 times more frequent than the PRO form; whereas, the PRO form was found to be 1.58 times more frequent in French studies (Baumann et al. 2014). As is common in RSA, we encode such grammatical (dis)preferences through assigning a higher **cost** to the dispreferred syntactic structure than to the preferred one. Thus, we propose that the (relevant) inventories of syntactic forms across the three languages are as follows:

Form (m)	German		English		French	
	$\llbracket m \rrbracket$	Cost(m)	$\llbracket m \rrbracket$	Cost(m)	$\llbracket m \rrbracket$	Cost(m)
Overt (<i>er/he/il</i>)	$\{p_S, p_O\}$	0	$\{p_S, p_O\}$	0	$\{p_S, p_O\}$	1.5
PRO			$\{p_S\}$	1	$\{p_S\}$	0

Following Arnold 2001, we assume that hearing a DP in subject position increases L’s expectation that this DP will serve as a referent in the subsequent discourse, which explains the cross-linguistic subject preference between sentences. We therefore take L’s beliefs after hearing (1), but prior to hearing (3), to be represented by the prior probability distribution $Pr(p_S) = 0.95; Pr(p_O) = 0.05$. We then apply the RSA iterated solution concept to this architecture (with soft-max temperature parameter = 1), and generate the predicted probabilities of subject interpretations, which mirror the experimental results.

LANGUAGE	BETWEEN SENTENCE	WITHIN SENTENCE
German	95%	95%
English	95%	72%
French	95%	15%

Thus, using these models, we show explicitly how cross-linguistic variation in pronoun resolution can be reduced to variation in the syntax of different languages.