Evaluating Bayesian Approaches to Pronoun Resolution Preferences: The Role of Presupposition

While the formal semantic and pragmatic underpinnings of intersentential pronoun resolution are comparably well understood, a comprehensive theory of the mechanisms that govern the psychological process of resolving a pronominal expression to an antecedent is still out of reach. This is arguably due to the large number of factors that are involved even in the most straightforward cases of a personal pronoun taking one of two potential antecedents from a preceding sentence.

Two recent theoretical approaches to anaphora resolution that attempt to make the contribution of a number of different factors as precise as possible are the Bayesian approach to pronoun interpretation presented by Kehler & Rohde, 2013; and the proposal combining verb and discourse semantics put forward by Bott & Solstad, 2014.

Kehler & Rohde propose to conceive of the probability to assign a referent to a pronoun (see the left hand term of (1), i.e. the interpretation bias to assign a pronoun a certain antecedent referent) as the solution to the following equation:

$$(1) \quad P(referent \mid pronoun) = \frac{P(pronoun \mid referent) \ \times P(referent)}{\sum\limits_{referent \in referents} P(pronoun \mid referent) \ \times P(referent)}$$

where the numerator terms denote the production bias to use a pronoun given a certain referent, and the bias to mention that referent, respectively. This account aligns (hearer-based) assignment preferences stemming from Hobbsian coherence relations with production preferences due to constraints of Centering Theory. In addition, it provides the ingredients to formulate precise predictions about the preferred interpretation assigned to pronoun given that the probabilities in the equation are specified.

In contrast, Bott & Solstad observe that a subclass of interpersonal verbs seem to carry a presupposition which influences the choice of preferred antecedent in discourses like the following:

(2) Peter rewarded John because he ... (e.g., ... won the competition).

According to Bott & Solstad, verbs like *to reward* carry the presupposition that, roughly, the referent of its direct object did something which is worth a reward. The authors argue that the preference to assign the pronoun to the referent of the object argument of the verb can be traced back to the presupposition triggered by the verb. The *because*-clause spells out content that is implied by the presupposition. The type of explanation that these presuppositional verbs engender make reference to external properties (attributed to NP2) rather than internal ones. Bott & Solstad give a precise semantic account of how this presupposition is derived from the semantics of the verb.

Both accounts have the merit of making explicit their assumptions about the mechanisms driving pronoun resolution, which are otherwise often left vague. The attempt to combine aspects of the two approaches, however, raises a problem which we think both of them need to address: **where should the contribution of the presupposed content be represented in the equation in (1)?** Should it enter into the computation of the probability of the left-hand term, i.e. the interpretation bias for a referent given a pronoun? This would predict that if a referent of a (given) pronoun satisfies the presupposition ascribed to it by the matrix verb, then this referent will become a more likely antecedent for the pronoun in question as compared to a referent that does not satisfy that presupposition. This, however, would be rather *ad hoc* and certainly not in line with the

reasoning behind (1), which attempts to derive this bias from the interaction of a number of conditionalized probabilities. Kehler and Rohde would attribute the influence of semantic factors like the content of the presupposition to the second multiplicand in the numerator of (1), i.e. the next mention bias, which is also modulated by e.g. coherence relations. But it is by no means clear whether the influence of presupposed content should directly affect the probability of a referent to be mentioned next.

To approach these question, we performed a sentence completion experiment. Specifically, we addressed the question whether it is possible to isolate the influence of the presupposition exerted by a verb in the matrix clause, in terms of antecedent bias, reference to the presupposition in the completion, and type of explanation. To this end, we compared pairs of verbs in which one verb carried a presupposition and the other did not, while keeping all other factors that enter into (1) equal. (3) gives an example of an experimental item (translated to English from the German original).

(3) a. Peter rewarded Paul with a bunch of flowers because he ...
    b. Peter gave Paul a bunch of flowers because he ...

16 participants completed 12 fragments of the type shown in (3), assigned to them in the two conditions by means of a latin square design, and interspersed with fillers. We predicted (i) that the (a)-conditions show a higher proportion of continuations in which 'he' refers back to the object of the matrix verb (i.e. Paul) than in the (b)-conditions, and (ii) that the continuations make reference to the presupposed content (what Bott & Solstad call the "occasioning state of affairs") more often in the (a) than in the (b)- condition. Our predictions were borne out: presuppositional verbs (condition(a)) engendered significantly more NP2 continuations (73%), and more continuations making reference to the presupposed content (80%) than their non- presuppositional counterparts (condition (b), 35% and 0%, resp.; both ps from a linear mixed model logistic regression $< .01$). In addition, our data showed a significant prevalence of external over internal reasons (in the sense of Bott & Solstad) in condition (a), and vice versa for condition (b) ($p < .05$).

We take this to be strong evidence that the influence of the presupposed content of verbs on pronoun interpretation can indeed be isolated. Verbs carrying a presupposition induce a strong NP2 bias for the pronoun, and their presupposition clearly affects the content and the explanation type of the completions produced by our participants.
This experimental result highlights the importance of the question of where in the equation (1) the effect of presupposition should be captured. Our results leave open the possibility that the right multiplicand in the numerator of (1)– -the probability to mention a referent next—might be a good candidate. But we suspect that the *semantic* nature of the effect cuts across the interpretation/production divide that is constitutive of Kehler & Rohde's Bayesian approach. For the time being, it remains to be shown how this approach can account for the influence of presuppositions.

References
Bott, O., & Solstad, T. (2014). From verbs to discourse – a novel account of implicit causality. To appear in: Hemforth, B., Schmiedtova, B., & Fabricius- Hansen, C. (eds.): *Meaning and Understanding across Languages*. Berlin, New York: Springer.
Kehler, A., & Rohde, H. (2013). A Probabilistic Reconciliation of Coherence- Driven and Centering- Driven Theories of Pronoun Interpretation. *Theoretical Linguistics*, 39: 1-37.