Helen de Hoop
Incremental optimization of pronoun interpretation

Helen de Hoop: Department of Linguistics, Radboud University Nijmegen, P.O. Box 9103, 6500 HD Nijmegen, The Netherlands. E-mail: h.dehoop@let.ru.nl

1 Introduction

All languages have personal pronouns, but the pronominal systems of the world mark different distinctions. While *person* is the basic feature of personal pronouns in the sense that all languages have different forms for first, second, and third person, languages differ greatly in other distinctions that they make. Semantic dimensions that might but need not play a role in the pronominal inventory of languages are number, gender, grammatical function, politeness, animacy, and even ethnicity (cf. Siewierska 2004, Heine & Song 2010, de Schepper 2013). The pronouns that Kehler and Rohde (target article, this issue) discuss are restricted to English third person singular masculine and feminine pronouns with the grammatical function of either subject or object, i.e., *he*, *him*, *she*, and *her*. These English pronouns are only used for animate, mostly human, individuals. Whereas for example in Dutch, *hij* 'he' and *hem* 'him' can also refer to inanimate things, this is impossible in English. Therefore, the pronouns examined by Kehler and Rohde only refer to people. Still, whenever there are two potential masculine antecedents around, such as *Mitt* and *Rick*, an anaphoric pronoun such as *he* is in principle ambiguous: it can refer back to either of the two male individuals, Mitt or Rick.

2 Optimal interpretations

Since people hardly ever make mistakes in the interpretation of pronouns, a mechanism must exist that leads them to the intended interpretation of a pronoun in context. I believe that this mechanism is optimization of interpretation (cf. Hendriks & de Hoop 2001). In Optimality Theoretic (OT) semantics, the input to the process of optimization is a form (an utterance) and the output an interpretation. Although in principle there is always an infinite number of interpretations possible, only one will come out as the optimal one, that is, the winner

(cf. Smolensky and Legendre 2006). It is possible to make a list of possible interpretations and then evaluate them against a set of ranked constraints, but this is just a way of modelling the process of optimization. The process itself is an automatic and parallel process in which the input (a pattern of activation) on the basis of the stored constraints (the weighted connections) leads to a pattern of activation (the output) with maximal harmony (Smolensky & Legendre 2006). So, upon hearing *Mitt narrowly defeated Rick, and the press promptly followed him to the next primary state* (example (1a) of Kehler & Rohde) hearers do not make a list of possible referents for *him* and then apply constraints in order to find out which one is the intended one. Rather, the optimal interpretation of *him* will emerge automatically on the basis of the content and the structure of this sentence as well as on our world knowledge, which are all reflected in potentially conflicting constraints. For example, a syntactic constraint such as PARALLELISM, that requires an anaphoric expression and an antecedent to have parallel syntactic functions or parallel thematic roles, can be in conflict with a constraint that favours the establishment of a rhetorical relation such as Contrast (Hendriks & de Hoop 2001, de Hoop & de Swart 2004, Hendriks et al. 2010). The fact that many constraints apply simultaneously is not predicted to slow down the process of optimization (cf. Smolensky & Legendre 2006).

How do readers arrive at the optimal interpretation of a pronoun such as *him* when there are two potential antecedents available in the linguistic context? Kehler and Rohde describe a series of psycholinguistic experiments, and argue on the basis of these that pronoun interpretation is affected by probabilistic expectations about coherence relationships within the discourse on the one hand, and expectations about what entities will be mentioned next, on the other. They present a probabilistic model that is capable of explaining pronoun interpretation preferences incrementally. The arguments for their model come from the results of sentence completion studies. This is remarkable in itself, because a sentence completion task is a language production task, whereas the proposed model is a model of language interpretation. In Optimality Theory, language production and language interpretation are modelled as two directions of optimization: speakers optimize from meaning to form – they want to convey a certain message and look for the optimal form to express it in the context – whereas hearers optimize from form to meaning – they hear a certain form and look for the optimal interpretation of this form in the context.

Obviously, a sentence completion task is rather complex compared to producing and interpreting utterances in everyday language use. First, the subject in such an experiment fulfils the role of a hearer (reader) looking for the optimal interpretation of the incomplete sentence. Second, the hearer switches to the role of speaker (writer) in order to complete the sentence which has started as the ut-

terance of another speaker. Unlike in a natural discourse, at the time of speaking there is no message that the speaker wishes to convey to the hearer. That is, there is no input meaning in this type of experimental task. The task is to produce a sentence not on the basis of an underlying meaning, but on the basis of an incomplete utterance of another speaker. This is why indeed a sentence completion task may tell us something about interpretation as well, because the speaker first has to interpret another speaker's utterance before the utterance can be completed in a sensible way.

The discussion point I wish to raise here is that a probabilistic model as proposed by Kehler and Rohde cannot account for the fact that in natural language use, pronoun interpretation is not just a matter of (high) probability, since any hearer in any context will arrive at one optimal interpretation, and hearers' interpretations of a pronoun in context converge. That is, the system of optimization is so robust that people hardly ever arrive at an optimal interpretation of a certain pronoun that is not the intended one (which would result in miscommunication). Consider for example the first experiment Kehler and Rohde discuss, by Stevenson, Crawley and Kleinman (1994), who instructed experimental participants to complete passages like the following:

(1) John seized the comic from Bill. He . . . . .
(2) John passed the comic to Bill. He . . . . .

Subjects were asked to complete the sentences in (1) and (2) and then Stevenson et al. "coded the **intended interpretations** of the pronouns". For contexts like (1), "they found an overwhelming bias (84.6%) to **interpret** the pronoun to refer to the subject/Goal (i.e., *John*)" (bold is mine, HdH). Whose interpretations have been coded here such that it can be said that an overwhelming bias is found? The subjects only had to complete the sentences; they did not have to code their own interpretations of the pronoun *he*. After the completion of the sentences, the reference of the pronouns was coded by the experimenters. Clearly, the experimenters coded their own pronoun interpretations, and apparently, they were so certain about their interpretations that these can be called "intended interpretations", even though in both contexts (1) and (2) the pronoun *he* is sometimes interpreted as referring to John and sometimes to Bill. This shows that the interpretation of a pronoun is not just a matter of high probability, because the assumption is that the pronoun is interpreted correctly by the experimenters in 100% of the cases. The final interpretation, i.e., the interpretation that is assumed to be correct, is based on the completed sentence. So, if the sentence is not yet completed, there is still some variation in how the pronoun will get interpreted after the completion of the sentence. The authors compute the probability that a pronoun gets

a certain interpretation when the sentence is not yet completed, but they do not explain how it is possible that after completion of the sentence, the probability that a pronoun gets a certain interpretation is 1, independently of whether the final interpretation is the one that was earlier preferred or not. Apparently, the information given by the completed sentence can overrule a rather strong bias on the basis of the preceding context.

3 Incremental optimization

In the remainder of this commentary I would like to show how the process of incremental optimization of interpretation can be modelled (cf. de Hoop & Lamers 2006). I will model the results of a psycholinguistic experiment described by Cozijn et al. (2011) to show this.[1] The experiment is about sentences such as (3) and (4) below:

(3) John beat Pete at the tennis match, because he had played very well
(4) John beat Pete at the tennis match, because he had been ill all week

There are two potential antecedents to which the pronoun *he* can refer. Constraints such as Parallelism and Continuing topic would favour the pronoun to refer to the subject of the previous clause, *John* (cf. Hendriks & de Hoop 2001, de Hoop 2004). Note that the complementizer *because* induces the establishment of a rhetorical relation of Explanation between the main clause and the subordinate clause. Since the cause of a winning event is more likely to be attributed to the winner than to the loser of the match, there is again a preference for *John* to be the antecedent of *he* (Cozijn et al. 2011). Verbs such as *beat* induce a preference for the attribution of the cause of the event to their subject, the agent of *beat*, whereas verbs like *criticize* have the opposite tendency, namely to attribute the cause to their object, the patient of *criticize* (Garvey & Caramazza 1974). Cozijn et al. (2011) show that implicit causality information is used early during processing and that it immediately contributes to the interpretation of the pronoun *he*. Their findings underscore that language processing is incremental and immediate.

Thus, in both (3) and (4) several constraints point into the direction of *John* to be interpreted as the antecedent of *he*. However, after finishing reading the complete sentence the pronoun *he* in (4) is interpreted as referring to Pete, be-

---

1 I am grateful to Laura Vugts and Miekske van der Sman for drawing my attention to this experiment and to the possibility of an incremental optimization of interpretation approach to the data.

cause our world knowledge dictates that being ill is a reason for losing a game rather than for winning it (although of course, more context could be added to make this other interpretation optimal). The incremental optimization of interpreting the pronouns in (3) and (4) above can be modelled by means of three potentially conflicting constraints. I will not use a constraint that induces the establishment of a causal relation because this would be satisfied by both relevant candidates due to the explicit use of the complementizer *because*. Also, since PARALLELISM and CONTINUING TOPIC yield similar results with respect to sentences such as (3) and (4) (they go hand in hand in choosing *John* as the antecedent of *he*), I only use CONTINUING TOPIC in my analysis. The second constraint that I use is called IMPLICIT CAUSALITY since it uses the lexical information of the verb (the *beat* type of verb versus the *criticize* type of verb) in order to determine the reference of the subject pronoun of the second sentence. Finally, the constraint WORLD KNOWLEDGE makes sure that the interpretation fits the expectations that we have on the basis of our world knowledge (such as having been ill a reason to lose but not to win a game). Since the optimal interpretation of *he* in sentence (4) is 'Pete', the constraint WORLD KNOWLEDGE outranks the other constraints. Based on the findings of Cozijn et al. (2011) we can also infer that the constraint IMPLICIT CAUSALITY outranks CONTINUING TOPIC. Now I can present the incremental optimization of interpretation approach to the sentences in (3) and (4). At the stage when the pronoun is encountered, the optimal interpretation of the pronoun is 'John':

Tableau 1: Incremental optimization of interpretation, stage 1, sentences (3) and (4)

John beat Pete at the tennis match, because he...	WORLD KNOWLEDGE	IMPLICIT CAUSALITY	CONTINUING TOPIC
☞ ‖he‖ = *John'*			
‖he‖ = *Pete'*		*	*

Subsequently, when the sentence is finished as in sentence (3) above, the optimal interpretation of stage 1 is maintained:

Tableau 2: Incremental optimization of interpretation, stage 2, sentence (3)

John beat Pete at the tennis match, because he...	... had played very well	WORLD KNOWLEDGE	IMPLICIT CAUSALITY	CONTINUING TOPIC
☞ ‖he‖ = *John'*	☞ ‖he‖ = *John'*			
‖he‖ = *Pete'*	‖he‖ = *Pete'*	*	*	*

When there is no congruity between the implicit causality of the verb and our world knowledge, the process of optimization in stage 2 yields a different winner, however, as shown in Tableau 3 below. This is called a 'jump' to another interpretation by de Hoop and Lamers (2006) who show that such a jump can also be reflected in certain ERP patterns.

Tableau 3: Incremental optimization of interpretation, stage 2, sentence (4)

John beat Pete at the tennis match, because he…	… had been ill all week	WORLD KNOWLEDGE	IMPLICIT CAUSALITY	CONTINUING TOPIC
☞ ‖he‖ = *John'*	‖he‖ = *John'*	⋆		
‖he‖ = *Pete'*	☞ ‖he‖ = *Pete'*		⋆	⋆

A similar jump to another interpretation arises in a sentence such as (5):

(5) John criticized Pete at the tennis match, because he was in a bad mood

When the pronoun is encountered, it is interpreted as 'Pete' in accordance with the object oriented implicit causality verb, which takes the patient of the verb to provide a plausible cause for the event of criticizing by the agent.

Tableau 4: Incremental optimization of interpretation, stage 1, sentence (5)

John criticized Pete at the tennis match, because he…	WORLD KNOWLEDGE	IMPLICIT CAUSALITY	CONTINUING TOPIC
‖he‖ = *John'*		⋆	
☞ ‖he‖ = *Pete'*			⋆

However, this interpretation turns out to be suboptimal when the sentence is completed as in (5). The resulting jump in interpretation is shown in Tableau 5:

Tableau 5: Incremental optimization of interpretation, stage 2, sentence (5)

John criticized Pete at the tennis match, because he…	… was in a bad mood	WORLD KNOWLEDGE	IMPLICIT CAUSALITY	CONTINUING TOPIC
‖he‖ = *John'*	☞ ‖he‖ = *John'*		⋆	
☞ ‖he‖ = *Pete'*	‖he‖ = *Pete'*	⋆		⋆

4 Conclusion

I hope to have shown that an incremental optimization of interpretation approach can account for the experimental data concerning pronoun resolution as well, since people can switch from one optimal interpretation to another at each stage of interpretation. Kehler and Rohde show that pronoun interpretation is a matter of probability as long as the sentence is not completed, because several continuations are possible, of which some are more likely than others. The fact that pronoun interpretation itself is not a matter of probability, however, follows from the fact that when the sentence is completed in a context, one interpretation remains as the optimal one.

References

Cozijn, Reinier, Edwin Commandeur, Wietske Vonk & Leo G.M. Noordman 2011. The time course of the use of implicit causality information in the processing of pronouns: A visual world paradigm study. *Journal of Memory and Language* 64, 81–40.

Garvey, Catherine & Alfonso Caramazza 1974. Implicit causality in verbs. *Linguistic Inquiry* 5, 459–464.

Heine, Bernd & Kyung-An Song 2010. On the genesis of personal pronouns: Some conceptual sources. *Language and Cognition* 2, 117–147.

Hendriks, Petra & Helen de Hoop 2001. Optimality theoretic semantics. *Linguistics and Philosophy* 24, 1–32.

Hendriks, Petra, Helen de Hoop, Irene Krämer, Henriëtte de Swart & Joost Zwarts 2010. *Conflicts in Interpretation*. Equinox, London.

de Hoop, Helen 2004. On the interpretation of stressed pronouns. In: Blutner, R. & H. Zeevat (eds.), *Optimality Theory and Pragmatics*. Palgrave/Macmillan, New York.

de Hoop, Helen & Monique Lamers. 2006. Incremental distinguishability of subject and object. In: Kulikov, L., A. Malchukov & P. de Swart (eds). *Case, Valency and Transitivity*. John Benjamins, Amsterdam/Philadelphia.

de Hoop, Helen & Peter de Swart. 2004. Contrast in discourse. *Journal of Semantics* 21, 87–93.

de Schepper, Kees. 2013. *You and me against the world? First, second and third person in the world's languages*. PhD thesis Radboud University Nijmegen. Utrecht: LOT Publications.

Siewierska, Anna. 2004. *Person*. Cambridge University Press, Cambridge.

Smolensky, Paul & Geraldine Legendre. 2006. *The Harmonic Mind. From neural computation to Optimality-Theoretic Grammar*. Cambridge, MA.