Towards a cognitively plausible computational model of reference

Jacolien van Rij (J.C.van.Rij@rug.nl)
Center for Language and Cognition Groningen, University of Groningen

Hedderik van Rijn (D.H.van.Rijn@rug.nl)
Department of Psychology, University of Groningen

Petra Hendriks (P.Hendriks@rug.nl)
Center for Language and Cognition Groningen, University of Groningen

Abstract

In this paper, we present a cognitive model of the production and comprehension of referring expressions in discourse. Our cognitive model incorporates the linguistic notion of accessibility and the mechanism of perspective taking to select the best referring expression to be used in a particular linguistic discourse. We discuss several predictions that follow from our implementation. On the basis of empirical evidence confirming these predictions, we argue that sufficient working memory (WM) capacity and sufficient processing speed are required for adult-like performance on the use of referring expressions in discourse.

Keywords: referring expressions; discourse; working memory; cognitive model; computational simulations.

Introduction

How do speakers determine the best referring expression to use in a particular situation? In many contexts, a speaker can choose between different forms that vary in the amount of information they convey. Generally, a pronoun (*he*) is used when reference is intended to the discourse topic, whereas more specific forms such as full noun phrases (*the pirate* or *a pirate*) or proper names (*Eric*) are used when reference is intended to less accessible referents or when new referents are introduced.

Several linguistic approaches have been put forward to explain how speakers select referring expressions. One influential approach assumes that the choice of referring expressions is crucially dependent on the properties of the discourse (a.o., Ariel, 1990; Givón, 1983): the speaker selects a referring expression that corresponds to the accessibility or saliency of that referent in the discourse. According to this view, the form of the referring expression signals the accessibility of the referent and thus helps the listener to determine the intended referent. Another approach assumes that a speaker needs to take into account the listener while selecting a form on an implicational scale of givenness (a.o., Gundel, Hedberg, & Zacharski, 1993; Hendriks, Englert, Wubs, & Hoeks, 2008). According to this view, referring expressions do not map directly onto an accessibility scale. Rather, different referring expressions are possible for referring to a referent that has a certain degree of accessibility. To accommodate the listener, the speaker should select a form that is as informative as necessary, but not more informative than that (cf. Grice's Maxim of Quantity, 1975). These two linguistic approaches often make the same predictions with respect to the choice of referring expressions, although they propose different mechanisms for how speakers determine their choice.

We implemented a computational model to investigate the production and comprehension of referring expressions. The computational model allows us to generate specific empirical predictions with respect to the way speakers determine the best referring expression to use in a particular discourse. Our model integrates the two linguistic approaches mentioned above. We argue that these two approaches are not necessarily incompatible, and that both accessibility and perspective taking are necessary when choosing an appropriate referring expression. In this paper, we present our model as a novel account for the choice of referring expressions and discuss three predictions that follow from our computational simulations.

Cognitive model

We implemented a computational model within the cognitive architecture ACT-R (Anderson, 2007) to simulate the production and comprehension of referring expressions. ACT-R provides a suitable framework to investigate the effects of discourse and cognitive factors such as working memory capacity and processing speed on linguistic reference, because ACT-R is a theory of human cognition with detailed assumptions about cognitive processes based on a range of experimental data. In production, the task of our model is to select the type of referring expression to be used, taking into account the preceding discourse and the listener, who should be able to recover the intended meaning. The model thus uses both accessibility and perspective taking to choose an appropriate referring expression.

Modeling accessibility

To produce a referring expression, the model first needs to know whether the intended referent is the discourse topic. The discourse topic is modeled as the referent with the highest accessibility, i.e., with the highest activation in declarative memory.

During on-line sentence processing the model builds a (simplified) representation of the preceding discourse: every referent in the discourse is represented in declarative

memory. Each representation (referred to as "chunk") has a certain amount of activation that reflects the accessibility of the referent in the current discourse. Within ACT-R, the activation of chunks is dependent on the frequency of use (the more frequently used, the higher the activation) and the recency of the last retrieval (the more recent the last retrieval, the higher the activation). The activation of chunks decays with time, but is increased when the chunk is retrieved. This activation mechanism implements the effect of the preceding discourse on the accessibility of a referent: the more frequently a referent is mentioned or the more recently the referent is referred to in the preceding discourse, the more accessible the referent is (see Arnold, 1998, for a review). In addition to this base-level activation, spreading activation can temporarily boost the activation of a chunk in a particular context, reflecting the usefulness of that chunk in that context. Chunks that are currently being processed spread activation to other, connected chunks in declarative memory. In our implementation, the subject of the previous sentence is a source of spreading activation, to reflect the observation that the subject of the previous utterance is likely to be the current discourse topic as well as the referent of a pronoun in the current utterance (Grosz, Weinstein, & Joshi, 1995; Stevenson, Crawley, & Kleinman, 1994). Due to the spreading activation from the subject of the previous sentence, the activation of the corresponding discourse referent becomes more activated in comparison with other discourse elements. As a result, the model will retrieve this discourse referent as the current discourse topic.

Within ACT-R, differences in spreading activation account for individual differences in working memory (WM) capacity, as the amount of spreading activation determines the ability to maintain goal-relevant information (Daily, Lovett, & Reder, 2001). This means that only if the amount of spreading activation is high, the chunk representing the subject spreads a large amount of activation and grammatical function is used in determining the discourse topic. On the other hand, if the subject spreads a small amount of activation, reflecting a low WM capacity, there will be no effect on the discourse elements associated with the subject. In that case, the effects of frequency and recency will be the main determinants of the discourse topic.

Thus, in our cognitive model, the discourse topic is not only determined by the preceding discourse, but also by WM capacity.

Modeling perspective taking

After determining whether the intended referent is the discourse topic, the model chooses a referring expression. The model then checks, by a step of comprehension following the first step of production, whether this referring expression will yield the intended referent for a listener. If the selected expression leads to a different referent than the intended one, the model will select another expression.

Two constraints guide the production and interpretation of referring subjects in the model (cf. Hendriks et al., 2008). The most important constraint, based on principles of economy, is a preference for using pronouns over more specific forms, such as full noun phrases. The second constraint requires a pronoun to be interpreted as referring to the discourse topic. Consequently, if the model intends to refer to a referent that is *not* the discourse topic, initially the model selects a pronoun, because it prefers to use a pronoun over a more specific form. However, in the second step of comprehension, taking the listener's perspective into account, the model interprets the pronoun as referring to the discourse topic. Because using a pronoun will result in an incorrect interpretation for the listener, the model then discards the pronoun and selects a more explicit full noun phrase instead.

Importantly, the extra comprehension step takes additional time. However, as production needs to be sufficiently fluent, performing both processes during on-line production requires sufficient processing speed (cf. Van Rij, Van Rijn, & Hendriks, 2010). Initially, the model cannot complete both processes, because this takes too much time. If the model is only able to complete the first process and does not take into account the listener's perspective, the model will produce pronouns all the time, even for referents that are not the discourse topic. In ACT-R, processes gradually become more efficient if they are used frequently (Taatgen & Anderson, 2002). Due to this learning mechanism of ACT-R, the model gradually acquires more processing speed and hence will be more likely to take into account the listener's perspective. When the model is able to take into account the listener's perspective, it will only use pronouns for reference to the discourse topic, and use full noun phrases for other discourse referents.

This cognitive model provides a novel explanation of how adult speakers determine the best referring expression in a particular discourse. This explanation combines accessibility and perspective taking: accessibility determines the discourse topic, which is necessary for selecting an interpretable referring expression in a particular context; perspective taking is used to evaluate whether the intended meaning is recoverable for the listener on the basis of the selected form.

On the basis of simulations of the cognitive model, we can evaluate the implications of these assumptions for the production and comprehension of referring expressions in discourse. In the following section, we discuss three specific and testable predictions that follow from our model.

Explanations and predictions

The predictions discussed in this section address the effects of discourse and cognitive factors such as working memory capacity and processing speed on the choice of referring expressions. The first prediction of our model pertains to children, who have a relatively low WM capacity and whose processing speed is still slow. The second prediction and

third prediction are concerned with adults in various WM load conditions.

Explaining underspecification

Children up to the age of 7 show underspecification in their production of referring expressions. They prefer to use pronouns over full NPs, even for reference to discourse elements that are not the discourse topic (Karmiloff-Smith, 1981; Koster, Hoeks, & Hendriks, in press). Such pronouns may cause misunderstanding for a listener, because pronouns tend to be interpreted as referring to the discourse topic. For this reason, adults generally use full noun phrases when referring to a referent that is not the discourse topic. In relation to children's production of underspecified pronouns, Koster et al. (in press) found that children who produce these unrecoverable pronouns tend to have lower scores on a WM task. Our cognitive model provides an explanation for these observations.

We compared simulations of our model with the data of Koster et al. (see Figure 1) (Van Rij, Van Rijn, & Hendriks, submitted). In these simulations, the model was presented with short stories about two referents of the same gender. Halfway through each story, the topic was shifted from the first referent to the second referent. This was done by changing the grammatical roles of the two referents, so that the second referent was the subject of the final sentence. After processing the presented sentences and building a representation of the discourse in declarative memory, the task of the model was to select a referring expression for the first referent, which was not the current discourse topic anymore.

Figure 1 shows the performance of the model with a low WM capacity, reflecting children's performance, and the performance of the model with a high WM capacity, reflecting adults' performance on the same task. The model's overuse of pronouns is caused by two different mechanisms: a low WM capacity and insufficient speed of processing. Our model predicts that low WM capacity will result in more underspecified forms than high WM capacity, because the low WM model is not very accurate in determining the discourse topic. With a high WM capacity, the model selects the subject of the previous sentence, which is the second referent, as the current discourse topic. With a low WM capacity, on the other hand, the model will only rely on frequency and recency of mentioning when determining the discourse topic. Therefore, the model will show a much-reduced preference for the subject of the previous sentence as the discourse topic, and will often choose the other referent. If the model (incorrectly) considers the first referent as the discourse topic, a pronoun is the preferred form to use. However, in this particular discourse situation, following a topic shift, a pronoun will result in misunderstanding for the listener.

Insufficient processing speed results in underspecified forms too, but for another reason, namely because the model is unable to check the recoverability of the chosen form. The model has a general preference for using a pronoun. Only if the model takes into account the listener's perspective will it decide to use a more specific form for referring to a referent that is not the discourse topic. Notice that low WM capacity does not prevent the model from gradually acquiring higher processing speed with linguistic experience.

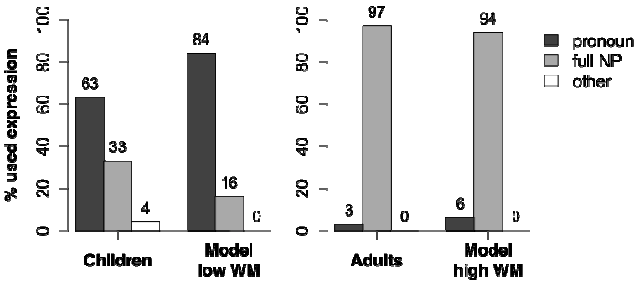

Figure 1. Production of referring subjects. The performance of the child and adult participants in the experiment of Koster et al. (in press) is compared with the performance of our low-WM and high-WM models (mean of 20 simulations) (from Van Rij et al., submitted).

Thus, on the basis of these simulations we argue that both sufficient WM capacity (which increases through maturation) and sufficient processing speed (which increases through linguistic experience) are important for adult-like pronoun use. Based on this explanation for children's use of underspecified forms, we predict that elderly adults with a low WM capacity but possibly still sufficient processing speed will have partly similar problems as children. As their low WM capacity prevents them from using grammatical information of the previous sentence in determining the discourse topic, they are expected to frequently use a pronoun for reference to a referent which they incorrectly take to be the discourse topic. Indeed, elderly adults with a mean age of 82 who scored significantly lower on a WM test than young adults produced pronouns for reference to a non-topic in a third of the cases, whereas the young adults hardly ever did so (Hendriks et al., 2008).

Predicting child-like performance in adults

A new empirical prediction following from our model is that adult listeners will show difficulty comprehending a topic shift if their WM capacity is limited. For example, if their WM is taxed by another task, they will be less likely to use the grammatical function of the referents in the discourse to determine the discourse topic. Rather, they will solely rely on the frequency and recency of the referents.

This prediction follows from the spreading-activation mechanism in ACT-R: goal-relevant information spreads a proportion of the total spreading activation to other chunks in declarative memory. If the number of sources from which activation is spread increases, the amount of spreading activation that is received by individual chunks decreases as the total spreading activation is fixed. In a situation of high

WM load, more information needs to be maintained in an activated state. Thus, more sources spread the fixed amount of spreading activation. As a result, the subject of the previous sentence spreads less activation to the discourse referents associated with the subject, and frequency and recency become more important in determining the current discourse topic.

This process will affect the interpretation of a topic shift, as illustrated in Table 1. In the first story in Table 1, Eric is the initial discourse topic but the topic shifts to Philip in the second half of the story. In the second story, Eric is the discourse topic throughout the story. The model predicts an effect of higher WM load for stories with a topic shift: at the beginning of sentence 4, adult readers will be less likely to choose the subject of the previous sentence, Philip, as the discourse topic, as the higher WM load reduces the additional spreading activation to Philip. Therefore, the model predicts that adults in a high WM load condition will more often choose the most frequently mentioned referent, Eric. On the other hand, the model predicts no effect of WM load for stories without a topic shift: at the beginning of sentence 4, adults prefer the most frequently mentioned referent, Eric, as the discourse topic, irrespective of WM load.

Table 1. Story with and without a topic shift in Dutch (with English translation) (from Van Rij et al., 2011).

Story with topic shift (+TS)
1. Eric/gaat/voetballen/in de sporthal. 'Eric is going to play soccer in the sports hall.'
2. Philip/vraagt/Eric/om mee te rijden/naar de training. 'Philip asks Eric to carpool to the training.'
3. Philip/haalt/Eric/na het eten/met de auto op. 'Philip picks up Eric after dinner by car.'
4. Hij/voetbalt/al twintig jaar. 'He has played soccer for twenty years.'
Story without topic shift (-TS)
1. Eric/gaat/voetballen/in de sporthal. 'Eric is going to play soccer in the sports hall.'
2. Eric/vraagt/Philip/om mee te rijden/naar de training. 'Eric asks Philip to carpool to the training.'
3. Eric/haalt/Philip/na het eten/met de auto op. 'Eric picks up Philip after dinner by car.'
4. Hij/voetbalt/al twintig jaar. 'He has played soccer for twenty years.'
Wie voetbalt al twintig jaar? 'Who has played soccer for twenty years?'

Using a dual-task experiment, we investigated the effect of additional WM load on the interpretation of pronouns in different discourse contexts as illustrated in Table 1 (Van Rij, Van Rijn, & Hendriks, 2011). Participants were asked to perform two tasks at the same time. Each trial, participants had to memorize either three (low WM load condition) or six (high WM load condition) digits. While memorizing these digits, they had to read stories of four sentences using the moving-window paradigm (Just, Carpenter, & Woolley, 1982), which were followed by a comprehension question (see Table 1 for examples). The stories featured two characters of the same gender, which were referred to with proper names. The final sentence started with a potentially ambiguous subject pronoun that could in principle refer to both characters. The comprehension question asked to name the referent of the ambiguous pronoun. The stories with and without topic shift only differed in the grammatical roles of the referents. The stories were tested in both WM load conditions. After selecting one of the referents as an answer to the question, participants had to type in the digits that were presented at the beginning of the trial.

As predicted, we found that WM load affects adults' *interpretation* of subject pronouns: adults less often selected the subject of the previous sentence as the referent of the pronoun in the high WM load condition than in the low WM load condition, but rather selected the most frequently mentioned other referent. Also as predicted, this effect of WM load was limited to stories with a topic shift and did not affect stories without a topic shift (Van Rij et al., 2011).

Thus, adults performed more child-like under high WM load and more often selected the incorrect referent. The question arises whether and how WM load affects the *production* of referring expressions in different contexts. Simulations of our cognitive model suggest that WM load may cause overspecification in the production of referring expressions, as will be discussed in the following subsection.

Predicting overspecification

Another prediction, one that has not been tested yet, concerns the effect of WM load on adults' choice of referring expressions in different discourse contexts. In the previous subsection, we discussed our finding that adults who have less WM capacity available due to a high WM load are more likely to select a non-subject as the referent of a pronoun, because they experience difficulty in using grammatical information when determining the discourse topic. Following the same argument, we predict that WM load may cause adults to produce overly specific referring expressions in particular contexts and use a full noun phrase or a proper name to refer to the discourse topic, although a pronoun would have been sufficient and hence would be more appropriate.

Overspecification as a result of WM load is predicted to occur if the discourse topic has been shifted to a referent that was less frequent in the preceding discourse than the previous discourse topic. In a low WM load situation, the speaker will tend to use a pronoun to refer to the new discourse topic immediately following the topic shift (a.o., Grosz et al., 1995). However, in a higher WM load

situation, the speaker may continue to use a more specific expression to refer to the discourse topic, for example a full NP or a proper name. This is because the higher WM load causes the speaker to rely more on frequency and recency of mentioning than on grammatical information when determining the discourse topic. If grammatical information from the previous sentence is not accessible due to a high WM load, a speaker may (incorrectly) assume that the new discourse topic is not the discourse topic. This can happen when another referent is more frequently or more recently mentioned in the preceding discourse. As a result, the speaker may use a full noun phrase or proper name until the accessibility of the new discourse topic has increased by frequent mentioning or recency.

Crucially, such overspecification is only predicted for adults or children who have insufficient WM capacity available but possess sufficient processing speed to check for recoverability of the intended referent. Thus, overspecification is not predicted for children with insufficient processing speed, because these children show a general preference for pronouns, which may lead to underspecification (cf. Koster et al., in press). In the high WM load situation, the model predicts that a speaker *correctly* chooses a full noun phrase after taking into account the listener's perspective, after having *incorrectly* determined that the referent is not the discourse topic.

To summarize, the difficulties in using grammatical information to determine the discourse topic when less WM capacity is available may prevent speakers from using a pronoun for reference to the new topic after a topic shift, resulting in overspecification.

Discussion

We have implemented a cognitive model that provides a novel explanation of how adult speakers determine the best referring expression to use in a particular context. The model integrates two influential linguistic accounts of the choice of referring expressions: based on the assumption that perspective taking is necessary for producing referring expressions that are interpretable for the listener (cf. Gundel et al., 1993; Hendriks et al., 2008), the choice of referring expressions in the model is also crucially dependent on the properties of the discourse (cf. Ariel, 1990; Givón, 1983).

Importantly, our model does not contain an explicit mechanism to avoid ambiguity. Ambiguity is not a problem if it results in successful communication, that is, if speaker and hearer converge to the same interpretation of a potentially ambiguous form within a particular discourse context. Only when ambiguity results in non-recoverability, which is when the hearer is expected to select a different referent than the one intended by the speaker, will the use of an ambiguous form be blocked as a consequence of the model's perspective taking mechanism.

Our computational model is implemented in the cognitive architecture ACT-R, which constrains simulation models to ensure cognitive plausibility. The constraints imposed on the models are based on experimental data and define how information is processed, stored and retrieved within modules, and how information is communicated between modules (Anderson, 2007). Our implementation suggests that linguistic explanations of referential choice can be formulated in terms of more general cognitive principles and mechanisms. In addition, our implementation sheds more light on the way cognitive factors affect the choice of referring expressions in discourse. In particular, our model predicts the occurrence of underspecification when WM capacity and processing speed are both insufficient, and predicts the occurrence of overspecification and errors in pronoun interpretation when WM capacity is limited but processing speed is sufficiently high.

Acknowledgments

Petra Hendriks gratefully acknowledges NWO (grant no. 277-70-005) for financial support.

References

Anderson, J. R. (2007). *How Can the Human Mind Occur in the Physical Universe?* New York: Oxford University Press, USA.

Ariel, M. (1990). *Accessing Noun-Phrase Antecedents.* London: Routledge.

Arnold, J. E. (1998). *Reference Form and Discourse Patterns.* Unpublished Ph.D. thesis, Stanford University.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science, 25*(3), 315.

Givón, T. (1983). Topic and continuity in discourse: An introduction. In T. Givón (Ed.), *Topic continuity in discourse: A quantitative cross-language study* (pp. 1-42). Amsterdam: John Benjamins Publishing.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics* (Vol. 3: Speech acts, pp. 41-58). New York: Academic Press.

Grosz, B. J., Weinstein, S., & Joshi, A. K. (1995). Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics, 21*(2), 203-225.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*(2), 274-307.

Hendriks, P., Englert, C., Wubs, E., & Hoeks, J. (2008). Age differences in adults' use of referring expressions. *Journal of Logic, Language and Information, 17*(4), 443-466.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology, 111*(2), 228-238.

Karmiloff-Smith, A. (1981). The grammatical marking of thematic structure in the development of language production. In W. Deutsch (Ed.), *The Child's Construction of Language* (pp. 121-147). London: Academic Press.

Koster, C., Hoeks, J., & Hendriks, P. (in press). Comprehension and production of subject pronouns: Evidence for the asymmetry of grammar. In A. Grimm, A. Müller, C. Hamann & E. Ruigendijk (Eds.), *Production-comprehension asymmetries in child language*. Berlin: De Gruyter.

Stevenson, R. J., Crawley, R. A., & Kleinman, D. (1994). Thematic roles, focus and the representation of events. *Language and Cognitive Processes, 9*(4), 519-548.

Taatgen, N. A., & Anderson, J. R. (2002). Why do children learn to say "Broke"? A model of learning the past tense without feedback. *Cognition, 86*(2), 123-155.

Van Rij, J., Van Rijn, H., & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language, 37*(3), 731-766.

Van Rij, J., Van Rijn, H., & Hendriks, P. (2011). WM load influences the interpretation of referring expressions *Proceedings of the 2nd workshop on Cognitive Modeling and Computational Linguistics* (pp. 67-75). Portland, OR: Association for Computational Linguistics.

Van Rij, J., Van Rijn, H., & Hendriks, P. (submitted). Production and comprehension of referring subjects: A computational model of the interaction between linguistic and cognitive constraints.