The object without qualities: referring with negative properties

Martijn Goudbeek (m.b.goudbeek@uvt.nl)
Tilburg center for Communication and Cognition, School of Humanities
Tilburg University, The Netherlands

Inge Haagmans (inge.haagmans@gmail.com)
Tilburg center for Communication and Cognition, School of Humanities
Tilburg University, The Netherlands

Emiel Krahmer (e.j.krahmer@uvt.nl)
Tilburg center for Communication and Cognition, School of Humanities
Tilburg University, The Netherlands

Abstract

Two experiments investigated the production and comprehension of referring expressions that contain a negative property ("the marker without a cap"). Experiment 1 showed that participants do use negative properties in their object descriptions, but that they were almost always accompanied by other properties, leading to referential overspecification. In experiment 2, participants identified objects based on descriptions that contained negative properties. While participants were faster in identifying objects that were described with preferred properties such as color, response times for objects described with a negative property ("the marker without a cap") and a positive property ("the marker with a cap") did not differ. The results provide behavioral grounds for extending referring expression generation algorithms to include negative properties.

Keywords: Referring expressions, speech production and comprehension, negative properties

(A sort of) Introduction

Image you are reading an interesting cognitive science manuscript and want to highlight an important passage. On the desk of your colleague (see figure 1) are two markers and since you prefer yellow markers, that is the one you would like her to pass on to you. You could phrase your request like this "Could you pass me the yellow marker, please?". In that case you have produced a referring expression with the property color as a means to distinguish between the two markers. However, a viable alternative to this question is "Could you pass me the marker without the cap, please?". In that case, you have used an *negative* property to refer to the marker of your choice. This paper investigates to what extent speakers produce referential expressions that contain negative properties and how listeners process these expressions. In doing so, we aim to inform computational models of referring expression generation.

Producing a suitable referring expression can be seen as a problem of choice (Krahmer & van Deemter, 2012). Which properties does a speaker include in the description when asking for the marker? In addition to color and the absence of a cap, the location ("the marker on the left") and size ("the slightly smaller marker") come to mind as possible distinguishing properties for your marker. How to choose between these properties, when they are all suitable candidates for inclusion in the referring expression? Many current computational approaches, such as Dale and Reiter's (1995) Incremental Algorithm use a fixed ordering of properties that are serially added to the description until all distractors have been ruled out. Since color is a property that is usually highly preferred (Pechmann, 1989), this is the first property that the Incremental Algorithm would add. Since the inclusion of color in the description uniquely identifies the item of your choice, the algorithm stops and produces "The yellow marker" and because the preference order is a fixed order, negative properties like "without a cap", would never be included in a referring expression.

Figure 1: The two markers on the desk of your colleague.

Standard REG (Referring Expression Generation) algorithms such as the Incremental Algorithm do not consider boolean operations such as negation in the generation process (Krahmer & van Deemter, 2012). Recently however, several attempts have been made to incorporate negation (and other boolean operators) in REG algorithms, either by extending the incremental algorithm (with a specific focus on referring to sets of objects) with boolean expressions (van Deemter, 2002; van Deemter & Krahmer, 2006), or by reinterpreting the problem of referring expression generation in terms of description logic or conceptual graphs (Areces et al., 2008; Croitoru & van Deemter, 2007). All of these approaches are computational in nature and until now the question of if, when, and how human speakers produce referring expressions with negative attributes has not been addressed. The behavioral data presented here can help inform computational approaches to referring expressions by, for example,

making the generated expressions more natural (Viethen & Dale, 2006; Dale & Viethen, 2010) or by providing inspiration for the further development of the algorithms. For example, the boolean extensions to the Incremental Algorithm proposed by van Deemter (2002) assumes that negative properties are less preferred than their positive counterparts. Here, we explicitly test this assumption with stimuli that do afford a description with negative properties, but can also be uniquely identified with other, more preferred properties.

While referring expression research is primarily concerned with the production of referring expressions, there is an increasing need to assess how listeners process the descriptions that are generated by REG algorithms (Krahmer & van Deemter, 2012). The most important criterion for a successful algorithm is whether the expressions generated mimic those of humans. However, humans might not always be good at taking the needs of their listeners into account (Horton & Keysar, 1996) and the references produced by human speakers might not be the most optimal ones. Thus, if our production experiment shows that speakers do use negative properties in their descriptions, this does not necessary mean that listeners will easily deal with such expressions. By combining a production experiment with a comprehension task, these issues will be addressed.

Two separate experiments will investigate the production and comprehension of referring expressions that contain an negative property. In the production experiment, three research questions are addressed. The first is whether speakers will produce referring expressions with negative properties at all in situations that afford (but not necessitate) the use of a negative property in a description. The second question is whether the number of positive properties necessary for a uniquely identifying description matters. It might be the case that speakers are more likely to use a negative property when the alternative means using a more complex description with, for example, two positive properties. This finding would contrast with the expressions generated by the Incremental Algorithm, which has no backtracking ability to take the length of the resulting referring expressions into account (Dale & Reiter, 1995; van Deemter, 2002). Finally, the phenomenon of overspecification is addressed. Speakers often produce referential expressions that contain more information than strictly necessary (for example, by referring to the marker with "the yellow marker without a cap"). Speakers have been shown to be more likely to produce overspecified references when they use dispreferred properties such as orientation (Goudbeek & Krahmer, 2012) and when they refer to target in more complex stimulus arrangements (Koolen et al., in press). Referring expressions with negative properties are arguably more complex and less preferred, leading speakers to overspecify descriptions that contain a negative property.

The comprehension experiment focuses on the processing of negative properties and addresses the question whether identifying objects that are described with negative properties takes more time than identifying objects that are described with positive properties.

Experiment 1: Producing referring expressions with negative expressions

In Experiment 1 participants produced descriptions of everyday objects. They could either refer to these objects with one or two *positive* properties ("the large marker" or "the large yellow marker") or with a *negative* property ("the marker without a cap")[1]. Additionally, this experiment investigated whether the number of properties necessary in the alternative description influenced the referential choices of the speakers.

Method

Participants

Twenty undergraduate students (eleven females) from the participant pool of Tilburg University took part in exchange for partial course credit. They were all native speakers of Dutch and were between 18 and 25 years old.

Materials

In the production experiment, the stimuli consisted of 96 sets of three objects. The target object was always presented in the middle and was marked with a black rectangle. Of the 96 stimuli, 64 were target objects that needed one of more properties for unique identification and and 32 were type-identifiable objects. Crucially, of the target objects, 32 could be described with a *negative* property such as "the marker without a cap" and 32 could be described with a *positive* property such as "the marker with a cap". The objects that had either positive or negative properties were a marker (with or without a cap), a cup (with or without a handle), a basket (with or without a lid) and a bottle (with or without a cap). See figure 2 for an example of two stimuli.

In addition to these properties, the target objects could alternatively be described with properties such as color or size. These are considered to be preferred properties in REG research (Dale & Reiter, 1995; Pechmann, 1989) and should thus serve as viable alternatives. To investigate whether the number of preferred properties necessary for identification plays a role in determining whether speakers will use a negative property, there was a condition where *one* positive property would suffice (e.g., "the orange marker", see figure 2a) and a condition where *two* positive properties were necessary (e.g., "the large orange marker, see figure 2b). In both conditions, one negative property ("the marker without a cap") would always suffice (see figure 2).

In addition, the experiment contained 32 type-identifiable stimuli that could be described by using type only (e.g., "the rabbit", "the strawberry"), leading tot a total of 96 stimuli.

Procedure

Participants were seated in a sound-attenuated room and were instructed to describe the object in such a way that a naive lis-

[1]Here and elsewhere we give English versions of Dutch originals.

(a) A target that can be referred to with *one* positive property ("the orange marker") or a negative property ("the marker without a cap").

(b) A target that can only be referred to with *two* positive properties ("the large yellow marker") or a negative property ("the marker without a cap").

Figure 2: The target stimuli used in the production experiment

tener would be able to identity the target object from the other two. The target object was always presented in the middle, and marked by a black square (see figure 2). Each stimulus was presented for five seconds, during which the participant's description was recorded. After the presentation and recording, a new set of objects immediately appears on the screen. Stimulus presentation and response recording took place on a PC and was controlled through the open-source package Opensesame (Mathôt et al., 2012). The experiment lasted about ten minutes, after which participants were debriefed and thanked for their cooperation.

Results

The descriptions of the participants were annotated with respect to which property they used in their description (size, color, pattern, and whether their descriptions contained a negative property or not. We also annotated whether a description contained any redundant properties, to see whether the use of negative properties might cause speakers to overspecify. First, we investigated whether participants used the negative property in their referring expressions at all in the condition that afforded to do so (see the left side of figure 3).

While our prediction is that speakers would use the negative property in their descriptions, the algorithm proposed in van Deemter (2002) would never include negative properties

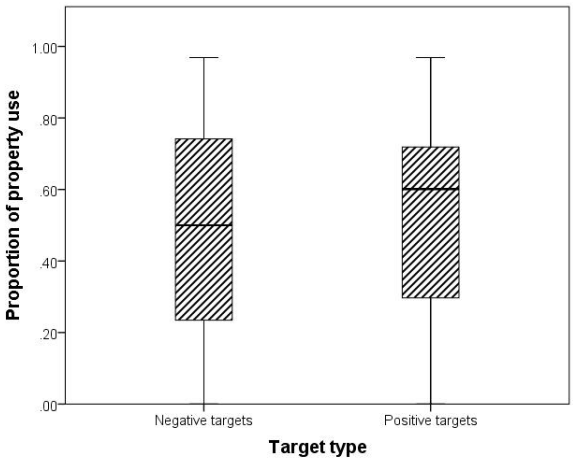

Figure 3: Proportion of descriptions with a negative property in conditions that afforded the use of negative properties (left) and proportion of descriptions with the corresponding positive property in conditions that afforded their use (right).

in the expressions it generates for these stimuli (since a preferred alternative is available). The results show that in almost half of the cases where the stimulus affords using a negative property, our participants did so (M = .47, SD = 0.31, 95 % CI = 0.33 - 0.62). A one sample t-test showed that this value did indeed differ significantly from zero (t (20) = 6.76, $p <$ .001, Cohen's d = 1.52). To investigate the effect of the target type (affording negative properties or not), we compared the proportion of descriptions with negative properties for stimuli that afforded the use of negative properties (M = .47, SD = 0.31, 95 % CI = 0.33 - 0.62) with descriptions with the corresponding positive properties for stimuli that afforded the use of positive properties (M = .52, SD = 0.30, 95 % CI = 0.38 - 0.66). The boxplot in figure 3 shows a large amount of overlap, indicating little difference between using a negative or a positive version of a property. A logistic regression with target type (positive versus negative) as outcome variable and proportion of properties used as predictor confirmed the lack of an effect of target type (β = 0.53, SE = 1.06, Wald = 0.25, $p = .62$, $R^2_{Nagelkerke}$ = 0.01).

Next, we tested the hypothesis that speakers would be more inclined to use negative properties when the alternative description required two positive properties. Figure 4 shows that speakers indeed produced more descriptions containing negative properties when the alternative contains two positive properties (M = .57, SD = 0.34, 95 % CI = 0.41 - 0.72) compared to when the alternative contains one positive property (M = .38, SD = 0.33, 95 % CI = 0.23 - 0.53). However, a logistic regression analysis with number (one versus two) as predictor and the proportion of descriptions with a negative property as outcome variable only yielded a marginally significant effect (β = 1.72, SE = 1.00, Wald = 2.98, $p < .08$, $R^2_{Nagelkerke}$ = 0.10).

Finally, we investigated whether referring with a negative

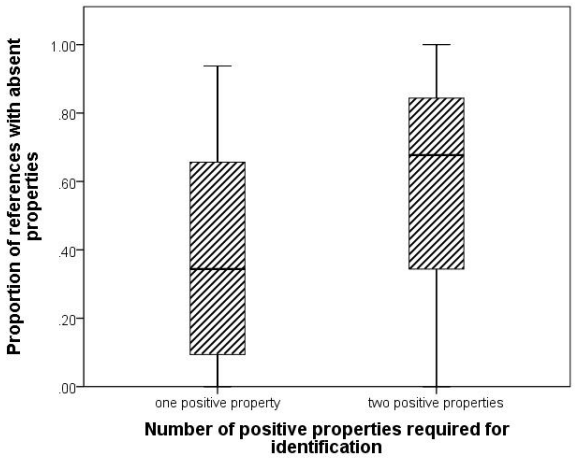

Figure 4: Proportion of referring expressions with a negative property in conditions with alternatives that required one or two positive properties for a uniquely identifying description.

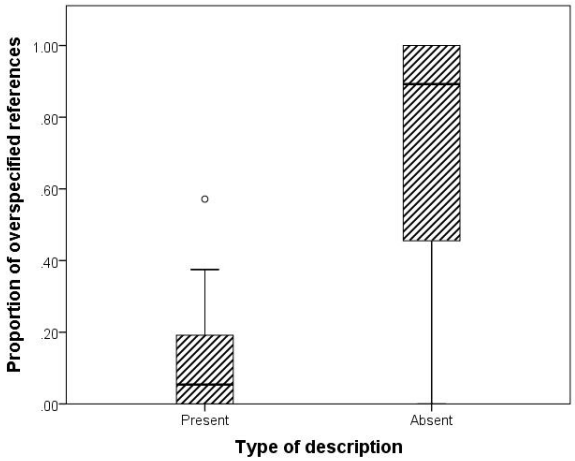

Figure 5: Proportion of overspecification in referring expressions that contained a negative property versus referring expressions that did not contain a negative property.

property causes speakers to overspecify (with a positive property) more than referring with positive properties only. For the subset of stimuli that afforded the use of negative properties, we calculated the proportions of overspecified references (defined as any reference that contains an additional property that would have been sufficent to uniquely identify it) for descriptions with negative and positive properties. Figure 5 clearly shows that when speakers use negative properties, they are more likely to use additional properties (M = .78, SD = 0.32, 95 % CI = 0.63 - 0.93) than when they do not use negative properties (M = .123, SD = 0.16, 95 % CI = 0.05 - 0.20). A logistic regression analysis with referring expression (negative versus positive) as predictor and the proportion of overspecified descriptions as outcome variable showed a significant effect (β = 2.62, SE = 1.75, Wald = 13.35, $p < .001$, $R^2_{Nagelkerke}$ = 0.73).

Discussion

The results of this experiment show that speakers certainly do not shy away from using negative properties in their referring expressions. In almost half of the cases participants included a negative property in their description when the stimulus afforded to do so. Importantly, our participants were never *forced* to use the negative property to identify the target referent: all objects could be uniquely identified by (a combination of) color, size or pattern, or by their type alone. This provides a psycholinguistic motivation for developing ways to generate referring expressions that contain negative attributes. Importantly, the boolean extensions of the Incremental Algorithm described in Areces et al. (2008), van Deemter & Krahmer (2006), and van Deemter (2002) do not fully do justice to the patterns observed here. For instance, while van Deemter (2002) assumes that negative properties are dispreferred, our speakers produced them just as much as their positive counterparts, even when properties that are considered to be more preferred (such as color and size) were at their disposal.

The comparison between alternatives that contained either one or two positive properties showed, albeit marginally significant, that the more complex the alternative expression becomes, the more likely speakers are to use a negative property in their description. This is difficult to explain for REG algorithms that depend on entering properties from a fixed preference order and that do not take into account the length or the complexity of the resulting referring expression. Interestingly, even though speakers often produce descriptions with negative properties, our analysis also showed that the resulting referring expressions hardly ever contain *only* these negative properties, but were often overspecified. This is in line with findings from previous studies such as Goudbeek & Krahmer (2012) and Koolen et al. (in press) that show that speakers are more likely to overspecify when their references include less preferred properties or when visual scenes get more complex. A possible explanation is that speakers could take the processing limits of their listeners into account (Arnold, 2008) and adjust the complexity of their utterances to suit.

Experiment 2: Understanding referring expressions with negative properties

Experiment 2 investigated the comprehension of referential expressions with negative properties.

Method

Participants

Twenty-eight undergraduate students (nineteen females) from the participant pool of Tilburg University took part in exchange for partial course credit. They were all native speakers of Dutch and were between 18 and 25 years old. None of the participants took part in Experiment 1.

Materials

For Experiment 2, the visual materials used in Experiment 1 were stripped from the black rectangle and were presented in the upper left, middle and upper right corner of the screen (see figure 6). They were complemented with a start box at the bottom of the screen. In addition, we recorded instructions to indicate the target object that contained five different ways to refer to the target. These referential expressions either used a positive property ("click on the marker with a cap"), a negative property ("click on the marker without a cap"), color or size ("click on the yellow marker"), color *and* size ("click on the large yellow marker"), and type-identifiable stimuli ("click on the strawberry"). All these descriptions were minimally specifying in that they provided sufficient, but not more, information to identify the target object. The instructions were spoken with a neutral intonation by a female speaker of Dutch that was unaware of the goal of the experiment. The position of the target was always either in the left or right upper corner and was counterbalanced across items. Since larger targets are easier to move towards and click on (Fitts, 1954), the size of the objects was counterbalanced as well (e.g., sometimes the instruction was "click on the small marker").

Procedure

Participants were seated in a sound-attenuated room and were given headphones to listen to the prerecorded instructions. They encountered the setup displayed in Figure 6 and could start the spoken description by moving their mouse pointer over the box labeled "START", after which response recording started. The participants' task was to click as quickly as possible on the object that was being described by the prerecorded referring expression. We used the software package MouseTracker (Freeman & Ambady, 2010) to present the images and speech stimuli and record the mouse movements and clicks. The experiment lasted about 20 minutes. After the experiment, the participants were debriefed and thanked for their cooperation.

Results

Table 1 shows the response times of the participants for the five different referring expressions. It should come as no surprise that the participants responded fastest to the type-identifiable items ("the strawberry"). Furthermore, the descriptions that used one were faster than those that used two two preferred properties. These in turn have faster response times than the conditions with either positive or negative properties, that do not seem to differ much.

These effects were evaluated statistically with a one-way within-subjects analysis of variance with type as a within factor with five levels (positive, negative, one, two, type-identifiable) and response time as dependent variable. This analysis showed a significant effect of type (F[4,108] = 187.01, $p < 0.001$, $\eta^2 = 0.87$). Planned contrast showed that response times to type-identifiable objects (M = 1.6, SD = 0.18) were faster than responses to descriptions with one preferred property (M = 1.72, SD = 0.26): F [1,27] = 38.46, $p < 0.001$, r = .76, which in turn were faster than responses to descriptions with two preferred properties (M = 1.72, SD = 0.26); F [1,27] = 117.81, $p < 0.001$, r = .90. These were faster than responses to descriptions with negative (M = 2.31 SD = 0.28) or positive properties (M = 2.39, SD = 0.26)); F [1,27] = 634.2, $p < 0.001$, r = .98), which did not significantly differ from each other (F [1,27] = 2.25, $p = 0.15$, r = .27).

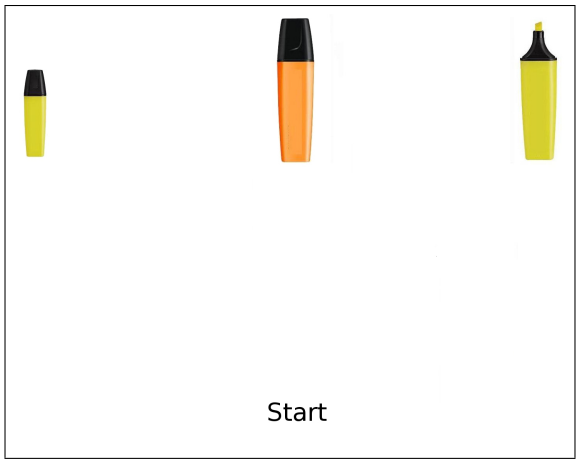

Figure 6: The stimulus presentation in the comprehension experiment. The target in the left upper corner can be described with "the marker without a cap" or "the large yellow marker".

Table 1: Response times, standard deviations and confidence intervals for the five different referring expressions in the comprehension experiment

Type	RT (s)	SD (s)	95% CI
Positive property	2.39	0.26	2.26 - 2.46
Negative property	2.31	0.28	2.18 - 2.40
Two properties	1.91	0.26	1.81 - 2.01
One property	1.72	0.24	1.63 - 1.82
Type identifiable	1.60	0.18	1.51 - 1.66

Discussion

In Experiment 2 participants identified objects based on to five types of referring expressions; these either contained a negative property ("the marker without a cap"), its positive counterpart ("the marker with a cap"), one preferred property ("the yellow marker"), two preferred properties ("the large yellow marker") or no properties at all (type-identifiable stimuli such as "the strawberry"). The results showed that listeners' response times closely follow the preferences of speakers. Our participants responded fastest to descriptions where using the targets type was sufficient for unique identification. They took (slightly) longer to respond to descriptions that contained preferred properties such as color or size, and it

took them significantly longer to identify targets that needed two properties to be uniquely identified instead of one. Not surprisingly, the response times were longest for the positive and negative properties that were not as preferred as size or color (having or lacking a cap, a lid, or a handle). Crucially, selecting the appropriate target that was described with a referring expression that contained negative properties did not take longer than selecting a target that was described with a positive property. This is in line with the observation from the production experiment, where speakers used the negative description ("the marker without a cap") as much as the the positive description ("the marker with a cap").

While we carefully controlled for the placement and size of the targets and their properties, the length and complexity of the descriptions was not the same for all descriptions. Descriptions that needed two properties contained more words than descriptions that needed only one property. However, the crucial comparison between descriptions with a positive or negative property differed in only one syllable ("with" versus "without"[2]). In addition, we measured response times from the start of the utterance, so our listeners could have already identified the target before the end of the referring expression, but see Arts (2004) for a discussion of measurements onsets in processing referential expressions. Although these issues might be difficult to control (referential expressions containing a negative property are inherently more complex than those with only one property), we do plan to take the length of the utterance into account in future work.

Conclusion

Taken together, the production and comprehension experiment provide experiential evidence for the use of negative properties in referring expressions. Speakers easily produce expressions such as "the marker without a cap" and listeners are not particularly troubled by them. These findings contribute to our understanding of speech production and perception processes in general, and provide important data for extending the scope of REG algorithms to descriptions containing negative attributes.

Acknowledgments

Many thanks to Carla van Gruijthuijsen for lending her voice to the stimulus material and to Sandra Augustin for her assistance in conducting the experiments. We gracefully acknowledge the financial support we received from The Netherlands Organization for Scientific Research (NWO, Vici grant 277-70-007 awarded to Emiel Krahmer).

References

Areces, C., Koller, A., & Striegnitz, K. (2008). Referring expressions as formulas of description logic. In *Proceedings of the fifth international natural language generation conference* (pp. 42–49). Stroudsburg, PA, USA: Association for Computational Linguistics.

Arnold, J. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, *23*(4), 495–527.

Arts, A. (2004). *Overspecification in instructive texts*. Tilburg University: Unpublished PhD thesis.

Croitoru, M., & van Deemter, K. (2007). A conceptual graph approach to the generation of referring expressions. In *Proceedings of the 20th international joint conference on artificial intelligence (ijcai)* (p. 2456-2461).

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive Science*, *19*(2), 233-263.

Dale, R., & Viethen, J. (2010). Empirical methods in natural language generation. In E. Krahmer & M. Theune (Eds.), (pp. 163–179). Springer.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, 381-391.

Freeman, J. B., & Ambady, N. (2010). Mousetracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behavioral Research Methods*, *42*(1), 226-241.

Goudbeek, M., & Krahmer, E. (2012). Alignment in interactive reference production: Content planning, modifier ordering and referential overspecification. *Topics in Cognitive Science*, *4*, 269-289.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*, 91–117.

Koolen, R., Goudbeek, M., & Krahmer, E. (in press). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*.

Krahmer, E., & van Deemter, K. (2012, March). Computational generation of referring expressions: A survey. *Computational Linguistics*, *38*(1), 173-218.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). Opensesame: An open-source, graphical experiment builder for the social sciences. *Behavioral Research Methods*, *44*, 314324.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*, *27*, 89-110.

van Deemter, K. (2002). Generating referring expressions: Boolean extensions of the incremental algorithm. *Computational Linguistics*, *28*(1), 37-52.

van Deemter, K., & Krahmer, E. (2006). Graphs and Booleans: On the generation of referring expressions. In H. Bunt & R. Muskens (Eds.), *Computing meaning* (pp. 17–53). Dordrecht, The Netherlands: Kluwer.

Viethen, J., & Dale, R. (2006, July). Algorithms for generating referring expressions: Do they do what people do? In *Proceedings of the 4th international conference on natural language generation* (pp. 63–70). Sydney, Australia.

[2]"with" translates to the monosyllabic "met" and "without" to the bi-syllabic "zonder" in Dutch