Factors causing overspecification in definite descriptions

Ruud Koolen [a,*], Albert Gatt [a,b], Martijn Goudbeek [a], Emiel Krahmer [a]

[a] Tilburg center for Cognition and Communication (TiCC), Faculty of Humanities, Tilburg University, P.O. Box 90153, NL-5000 LE Tilburg, The Netherlands
[b] Institute of Linguistics, University of Malta, Centre for Communication Technology 404A, Tall-Qroqq, Msida MSD 2080, Malta

A B S T R A C T

Speakers often overspecify their target descriptions and include more information than necessary for unique identification of the target referent. In the current paper, we study the production of definite target descriptions, and explore several factors that might influence the amount of information that is included in these descriptions. First, we present the results of a large-scale experiment investigating referential overspecification as a function of the properties of a target referent and the communicative setting. The results show that speakers (both in written and oral conditions) tend to provide more information when a target is plural rather than singular, and in domains where the speaker has more referential possibilities to describe the target. However, written and spoken referring expressions do not differ in terms of semantic redundancy. We conclude our paper by discussing the implications of our empirical findings for pragmatic theory and for language production models.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Speaking often involves referring to things in order to identify them. Speakers already learn how to refer from an early age (Ford and Olson, 1975; Yip and Matthews, 2007), and referring expressions can be found across all languages. When referring, language users constantly need to decide about which and how much information to provide in order to make communication successful. For example, consider the two persons in Fig. 1.

Suppose you want to point out person A (the *target*) to an addressee. To do this adequately requires describing person A in such a way that the addressee can distinguish him from person B (the *distractor*). It stands to reason that there is a choice between different distinguishing attributes in order to make person A identifiable; definite descriptions like "The man wearing glasses", "The man with the white shirt", and "The man without a tie" all contain sufficient information for the addressee to single out the target.

Although the aforementioned target descriptions contain different attributes to distinguish person A, they have in common that they are all minimal identifying descriptions of the target referent. That is, apart from the type, or basic category to which the target entity belongs (namely, 'person'), they contain just enough additional information (target attributes) for the addressee to single out the target. However, several psycholinguistic studies have revealed that speakers often overspecify their references and include more information than is strictly necessary for identification (e.g., Arts, 2004; Brennan and Clark, 1996; Engelhardt et al., 2006; Pechmann, 1989). Thus, when referring to person A in Fig. 1, a speaker

---

* Corresponding author. Tel.: +31 134663584; fax: +31 134662892.
  E-mail addresses: R.M.F.Koolen@uvt.nl (R. Koolen), albert.gatt@um.edu.mt (A. Gatt), M.B.Goudbeek@uvt.nl (M. Goudbeek), E.J.Krahmer@uvt.nl (E. Krahmer).
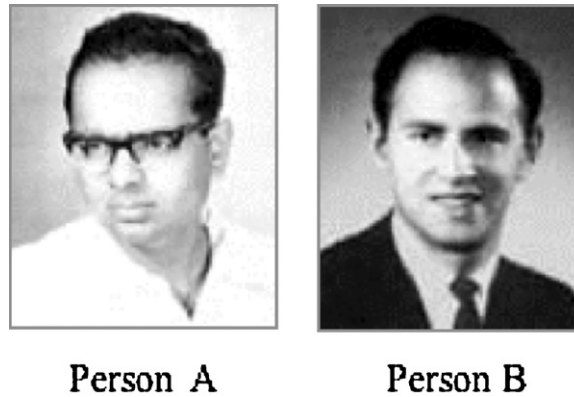
Fig. 1. Two possible target referents.

could produce an overspecified expression like "The black-haired man wearing glasses" instead of a minimally specified expression.

While the phenomenon of referential overspecification is by now fairly well established, it remains largely unexplored how factors such as properties of the target referent and properties of the communicative setting influence it. In this paper we therefore present the results of a large-scale language production experiment in order to investigate whether and to what extent this is the case.

2. The production of definite reference

When speakers refer, they usually aim to identify a particular target referent to a listener (e.g., Searle, 1969). How do speakers decide which attributes to include in a target description? A straightforward way to do this is by selecting those attributes that rule out other entities in the context of the target. Olson (1970) emphasized this primarily contrastive function of referring expressions by revealing that decisions regarding the semantic content of a referring expression (the attributes of a referent to include) are determined by the speaker's knowledge of the intended target referent and the surrounding distractor objects. This implies that referring expressions are contrastive by means of their distinguishing attributes: speakers identify a target referent by ascribing a combination of attributes to it that cannot be ascribed to the distractor objects (Bach, 1994). In everyday discourse, the attributes that speakers include in their referring expressions are determined by endophoric or exophoric considerations (or, often, by a combination of the two).

In case referring expressions are endophorically used, this means that the attributes they contain are determined on the basis of previously described objects in the unfolding discourse. Sometimes, however, the previous expressions in the discourse are not taken into account; the referring expressions are then exophorically used. When producing exophoric expressions, the speaker only uses the knowledge available in the direct physical context of the target (that is, the target itself and its immediate distractors) to determine which and how much information to mention. Since exophoric referring expressions do not take previous expressions into account, and especially serve to identify objects in an environment, they are considered more basic than endophoric ones.

Although referring expressions can occur in numerous forms, the focus in this paper is on definite noun phrases, rather than on pronouns, nominalizations or other linguistic mechanisms used for reference. The definite noun phrases under study here consist of a definite article, a head noun and possibly one or more modifiers (e.g., "The man wearing glasses" or "The red button"). These modifiers may be realised in different ways, for example adjectives ("red"), restrictive relative clauses ("which is red"), and so on. Furthermore, the noun phrases we study refer to physical objects in the world (such as persons or chairs), and not to abstract entities (such as "reality" or "democracy"). It needs to be emphasized that we take a purely exophoric perspective in studying reference: the expressions we study are first mention definite descriptions, and are solely intended to identify a target referent to an addressee without serving any other communicative goal. These kinds of definite descriptions are, among others, commonly used in written text (Poesio and Vieira, 1998), and in written and spoken instructive discourse (see for example, Maes et al., 2004).

2.1. The occurrence of referential overspecification

As the example used in the introduction of this paper shows, speakers can describe a target in many different ways, using different types and amounts of information. In this respect, several studies have shown that speakers often include more information in their descriptions than strictly necessary for unique target identification. Thus, instead of being minimally specified, referring expressions are often overspecified. For example, Deutsch and Pechmann (1982) designed a language production game in which speakers were presented with eight objects (in four domains) and were asked to

refer to one object that they would like as a birthday present. Once the addressee was able to identify the speaker's target, he placed it in a box. Deutsch and Pechmann found 28% of the adults' referring expressions to contain one or more redundant attributes of the target. In a language production experiment by Pechmann (1989) speakers were asked to uniquely identify one target object from a set of distractor objects (such as cars, chairs, etc.). The objects differed in type and/or colour and/or size. Pechmann found 21 percent of the speakers' expressions to contain at least one redundant attribute and thus to be overspecified. Nadig and Sedivy (2002) also found evidence for referential overspecification in an experiment in which speakers were asked to describe a target object (for example a big glass) in a display of four objects to an addressee. One contrasting distractor object of the same class as the target (for example a small glass) was either absent, visible only to the speaker, or visible to both speaker and addressee. Results revealed that when the contrasting target object was not visible to the addressee, over 50 percent of the descriptions contained redundant modifiers. Similar results were found by Maes et al. (2004), who performed an experiment in which participants produced written instructive texts on how to use a radio alarm. Their results showed both initial references (52%) and anaphoric referring expressions (75%) to be frequently overspecified. In Engelhardt et al. (2006), speakers produced referring expressions by asking them to instruct an addressee (who was a confederate) to move real target objects on or in some other object (e.g., an apple on a towel). In the *matching* location condition, the target object had to be moved from one location to another of the same type (e.g., from one towel to another towel), while in the *different* location condition the target object was moved to a different type of location (e.g., from a towel to a box). Overall, referring expressions contained redundant information approximately one third of the time.

2.2. Overspecification vs. the Gricean Maxim of Quantity

The occurrence of overspecification seems not to be in agreement with the Cooperative Principle formulated by Grice (1975), and with the Maxim of Quantity in particular. This conversational maxim consists of the following two components (Grice, 1975:45):

1. Make your contribution as informative as is required (for the current purposes of the exchange);
2. Do not make your contribution more informative than is required.

With respect to the amount of information that a speaker should provide to identify a target to an addressee, a strict interpretation of the Maxim of Quantity seems to suggest that the speaker should provide enough information to identify the target (Quantity 1), but not more information (Quantity 2). Hence, this would imply that any attribute that is not needed for target identification should be considered redundant, and that the inclusion of redundant attributes would lead to overspecified target descriptions.

Given that people are known to cooperate when they are in a conversation (Brennan and Clark, 1996; Clark and Wilkes-Gibbs, 1986), addressees will expect all information included in an expression to be relevant. Therefore, interpreting the Maxim of Quantity in a strict manner (such as described above) seems not to be in line with what Grice intended (Bach, 2006): one can think of other reasons why a speaker would want to include information that does not directly serve the target identification goal. As Grice states himself, the information level of a contribution should depend on the purpose that a speaker has in uttering it. Mooney (2004) further elaborates on this by arguing that the Maxims that are part of Grice's Cooperative Principle are generally contextually sensitive, which – with regard to the Maxim of Quantity – implies that it depends on the context in which an utterance takes place whether all information provided is relevant. According to Mooney, the notion of activity types by Levinson (1979) can be used to contextualize the Cooperative Principle, since these activity types include (non-linguistic) goals of discourse that involve more than the communication of information alone.

In any case, it is hard to give an a priori definition of what determines whether a given conversational contribution is overspecified in a given context. However, for reasons of making the Maxim of Quantity (and the phenomenon of overspecification in particular) testable in a controlled setting, there is a variety of papers in pragmatics (e.g., Arts et al., 2011; Engelhardt et al., 2006), psycholinguistics (e.g., Olson, 1970; Pechmann, 1989), and computational linguistics (e.g., Dale, 1989; Dale and Reiter, 1995; Jordan, 2000, 2002) that investigate overspecification in the light of target identification. As stated in the beginning of section 2, this is also what we aim to do in this paper, and we therefore regard descriptions as overspecified in case more information is included than strictly necessary for unambiguous identification of the target. In this way, we investigate several factors that may influence the specification level of definite target descriptions, aiming to uncover the cognitive processes that underlie the production of such descriptions with a view to explaining why speakers, even in apparently 'simple' settings, overspecify.

2.3. Why do speakers overspecify?

Why do speakers often provide more information than strictly necessary when describing a target referent? The traditional cognitive view on human referring behaviour emphasizes that speakers design their referring expressions in such a way that the addressee can efficiently identify the target (Arnold, 2008). This suggests that referring is an addressee-oriented process, and that one reason why speakers often overspecify their referring expressions could be that it helps the addressee to identify the target more quickly. This view is also compatible with some pragmatic accounts of reference,

notably Searle's (1969). Several empirical studies have provided support for this addressee-oriented view by revealing that listeners find it easier to identify a target referent when the speaker's description is overspecified, rather than minimally specified (Deutsch, 1976; Nadig and Sedivy, 2002; Paraboni et al., 2006; Sonnenschein, 1984; Sonnenschein and Whitehurst, 1982). Further support came from Engelhardt et al. (2006), who demonstrated that addressees judging the quality of instructions do not rate overspecified referring expressions to be any worse than minimally specified ones.

One possible explanation for overspecified references being beneficial for addressees could be that addressees create a *conceptual gestalt* of the target object for which they have to search (Levelt, 1989). Perceptually salient target attributes such as the basic-level category ('type') and 'colour' are central to a gestalt, because they help the addressee to construct a mental representation of the target. This is in line with studies by Eikmeyer and Ahlsèn (1996), Pechmann (1989), and Schriefers and Pechmann (1988), who found 'type' and 'colour' to be omnipresent in referring expressions, irrespective of their contrastive value. This would suggest that speakers redundantly mention such perceptually salient target attributes because they help listeners to complete their conceptual gestalt and speed up the identification process. This was confirmed by Arts et al. (2011), who revealed that in individual expressions, overspecified referring expressions that contained perceptual information (e.g., about shape, colour, or size) led to significantly shorter identification times than minimally specified expressions. Davies and Katsos (2009) have also emphasized this pragmatic contribution of perceptually salient attributes to successful communication.

However, there is also some evidence that overspecification is detrimental for listeners, at least beyond a certain point. In a separate experiment using eye tracking, Engelhardt et al. found that listeners tended to be slowed down by overspecified expressions during identification. This result has since been replicated using a different methodology by Engelhardt et al. (2010). In a related vein, eye tracking on reference resolution in the Visual World Paradigm (Tanenhaus et al., 1995) has consistently observed what is known as a *point of disambiguation effect*, whereby listeners process incoming information in a description incrementally as a description is heard, circumscribing the domain of possible referents to exclude distractors that do not have the attributes heard so far in the description until a point is reached where the referent has been identified (Eberhard et al., 1995; Sedivy et al., 1999). However, this seems to suggest that no further domain circumscription would occur beyond the point where a target has been identified, even if further information were available in the incoming description.

Previous research on reference production has shown that facilitating the identification process is not the only reason why speakers overspecify their referring expressions. Speaker-oriented processes may also play a role. For example, speakers often contrast a particular target referent with a previously mentioned one, despite the fact that this previous target referent is no longer visually present in the visual domain (Levelt, 1989). This kind of displaced communication (Spivey and Richardson, 2008) makes the speaker include the distinguishing attribute of the previous target object in the reference to the new target object, irrespective of its contrastive value. Further evidence for speaker-oriented processes comes from Belke and Meyer (2002), who conducted an eye tracking study to explore speakers' preferences for certain attributes. In their experiment, Belke and Meyer had participants judge whether certain target and distractor objects were similar or different. Results showed a high proportion of overspecified referring expressions, due to the speakers' tendencies to redundantly mention colour in their descriptions. Belke and Meyer claim that this result was due to the fact that speakers preferred perceptually salient attributes (such as 'colour'). Furthermore, because these attributes are absolute, they do not require the speaker to compare the target with the distractor object to determine the right value. In contrast, Belke and Meyer also found the opposite to be the case for attributes such as size, which do require a target referent to be compared to its distractors in order to determine the right value (e.g., whether the referent is large or small).

2.4. Computational implications of referential overspecification

The production of referring expressions has received considerable attention in Natural Language Generation (NLG), a subfield of Natural Language Processing and Artificial Intelligence (AI) that aims to build systems that automatically generate natural language text or speech from non-linguistic information (e.g., from a database; Reiter and Dale, 2000). Practical applications of NLG include, among others, the automatic generation of weather forecasts (Goldberg et al., 1994; Reiter et al., 2005), summarization of medical information (Gatt et al., 2009), and generation of instructions in contexts such as museum tours (Stock et al., 2007).

Given the ubiquity of referring expressions in natural language, it is no surprise that NLG systems typically also require algorithms that compute distinguishing descriptions for objects (Mellish et al., 2006). Various Referring Expression Generation (REG) algorithms have been proposed. As in the case of psycholinguistic research in this area, REG algorithms have typically taken the Gricean Maxim of Quantity as a starting point. For example, Dale's Full Brevity algorithm (Dale, 1989, 1992) was based on a strict interpretation of the Maxim, seeking to find the shortest possible description (in terms of the number of attributes included) for a target referent. This turned out to be computationally expensive, because it involved exhaustive search through the space of possible combinations of the target attributes in order of their length. A more relaxed interpretation of the Maxim is afforded by the Incremental Algorithm (Dale and Reiter, 1995), which was partly motivated by some of the psycholinguistic phenomena identified above, particularly the fact that some attributes are preferred over others. To do this, the Incremental Algorithm performs a "hillclimbing" search along a predetermined *preference order*, which lists attributes in order of preference (for example, placing 'colour' before 'size'). Given an intended referent, the algorithm searches along the preference order, adding an attribute to a description if it has some contrastive value and terminating

once the description is fully identified. Since the algorithm does not backtrack to remove attributes that turn out to be redundant, a description can be overspecified.

Since the work of Dale and Reiter (1995), there have been many further developments of the original Incremental Algorithm, for example to handle plural referring expressions, in particular, those which involve logical disjunction of attributes, and are typically realised as coordinate NPs like "the blue and brown books" (Van Deemter, 2002; Gatt, 2007). In addition, there has also been some work on developing frameworks that can accommodate different algorithms, such as graph-theoretic approaches (Krahmer et al., 2003).

The performance of REG algorithms is usually evaluated by comparing them to a corpus of human-authored *reference* descriptions, the idea being that the quality of an algorithm is reflected by the extent to which its output matches that of humans on the same input domains (see, among others, Viethen & Dale, 2006; Gupta and Stent, 2005; Gatt et al., 2007; Gatt and Belz, 2010). From a comprehension perspective, *effectiveness* is also considered a conceivable criterion of adequacy of an algorithm's output (Van Deemter et al., in press). Paraboni et al. (2007) argue that referring expressions can be regarded as effective when they are easy for an addressee to comprehend and/or resolve.

As we have observed, considerations of computational efficiency, as well as psycholinguistic considerations have led to a relaxation of the Gricean Maxim of Quantity in determining how REG algorithms perform their task. Still, there remain several open questions concerning when an algorithm should indeed overspecify, and how. Part of the reason for this is that psycholinguistic results have to date provided only a partial picture of the processes underlying referential communication. For example, while a preference for certain attributes that results in increased likelihood of overspecification has been established, it is not clear whether speakers are also influenced by the type of domain they are considering (e.g., whether they are talking about items such as chairs and tables, or completely different entities such as human beings). Nor is it clear to what extent overspecification is influenced by the nature of the target referent, in particular, whether both singular and plural references are equally likely to be overspecified. For example, Van Deemter's (2002) extension of the Incremental Algorithm to handle plurals was found to return highly complex expressions with a great degree of redundancy, leading some researchers to propose a return to strict Gricean principles for plurals (Gardent, 2002). Whether this is justified remains something of an open question (but see Gatt, 2007, for evidence that overspecification occurs in plurals as well).

Another issue that has not received much attention in the computational literature is the possible influence of different communicative settings on reference. Many of the algorithms discussed here are agnostic as to whether they are used in the context of an NLG system for text or speech, or whether the system is generating language in a situated setting or in a dialogue. There are some notable exceptions, such as Heeman and Hirst's (1995) computational interpretation of Clark and Wilkes-Gibbs's (1986) collaborative model for reference and, more recently, work in situated dialogue settings (Stoia et al., 2006; Kelleher and Kruijff, 2006; Byron et al., 2009). However, these computational approaches tend to have little to say about overspecification per se. An interesting exception is the work of Jordan (2000, 2002), whose investigation of redundancy in dialogues involving a joint task showed that interlocutors frequently produced descriptions that were overspecified in the context of dialogue, for instance, by repeating information that had already been used previously. Jordan argued that one reason for this could be that identification in the context of dialogues is often not the only goal of a referring expression. For example, a speaker may repeat information to signal agreement with an interlocutor. A computational implementation of this model showed that it matched human behaviour (as reflected in a dialogue corpus) better than some existing algorithms (Jordan and Walker, 2005). This effect of repetition is related to Pickering and Garrod's (2004) notion of *alignment*, which holds that the expressions that are produced earlier in the interaction influence the intonation patterns and syntactic structures of the ones that are produced later in the interaction. Goudbeek and Krahmer (2010) showed that alignment in referring expressions also occurs at the level of content selection (i.e. the target attributes that are included).

In summary, there are a number of open questions in pragmatics whose answers may help the development of better REG algorithms. One set of questions concerns the aspects of the domain that result in overspecification, and also how descriptions to singular and plural targets differ in terms of overspecification. Another concerns the difference between modality (speech vs. writing) and interactivity (monologue vs. dialogue settings). It is one of the motivations of the present study to contribute to answering these sets of questions.

2.5. Factors causing referential overspecification

In order to make REG algorithms described above perform better in terms of the amount of information that they include in generated output, some of the factors causing human speakers to produce overspecified references need to be further addressed and discussed. We therefore investigate two factors that we expect to cause referential overspecification: properties of the target referent and properties of the communicative setting.

2.5.1. Properties of the target referent

We hypothesize referential overspecification to be influenced by the properties of the target referent. We expect that targets that require more effort to refer to – in a sense to be made more precise below – will more often result in overspecified references than targets that are easy to refer to. This is in line with an observation by Van der Sluis and Krahmer (2007). In a study looking at the production of deictic (pointing) gestures, they found the difficulty of the referential task to influence overspecification: in difficult tasks (i.e. pointing to far away targets), speakers used more words and included more locative

relations in their descriptions compared to simple tasks. We investigate two kinds of properties of the target that may influence referential overspecification: the type of domain in which a target occurs, and the cardinality of the target (i.e. whether it consists of one referent or more than one).

For the *domain* in which a reference is produced, we hypothesize that expressions produced in domains where the objects vary more in terms of the number of potential attributes (and therefore afford speakers with more referential possibilities) will be more likely to contain more information than expressions produced in a domain where the speaker has fewer attributes to choose from. In order to investigate this, we made speakers refer to target objects in two different domains: one domain consisting of pictures of furniture items that vary in terms of a small number of attributes, and one domain in which entities are photographs of real people that differ on a higher number of attributes. We expect references to targets in the latter domain to be more frequently overspecified. By studying this, we aim to find out whether the specification level of target descriptions is affected by the referential possibilities that speakers have in a given domain, and thus, how different domains might lead to different specification levels. Clearly, this is not the only factor that determines the 'complexity' of a domain. For example, world knowledge will have an impact on what speakers choose to say about an entity. However, our hypothesis targets the specific issue of the amount of choice available to a speaker and the impact that this has on the ultimate decision of how to identify an object.

For *cardinality* (i.e. whether references are singular or plural), we hypothesize that plural references are more likely to be overspecified than singular references. More specifically, we expect that speakers are more likely to include redundant attributes in their references to two target objects than in their references to one target object. The reasoning behind this is twofold. First, there may be competition between the two target objects in plural references, which might force speakers to divide attention. As a result, the referring task could become more difficult, since dividing attention may lower the activation level that each of the two targets has in the speaker's mind. This explanation is in part based on findings by Arnold and Griffin (2007), who found evidence for an effect of cardinality on referential overspecification, albeit in a different context from ours, looking at when speakers use a pronoun. They report the results of two experiments demonstrating that speakers are less likely to use a pronoun when they need to describe two characters in a cartoon than when they need to describe a single character. In the present case, we expect that the effect of competition and divided attention may result in less control over the amount of information given about an entity in a definite description. A second reason why we expect plural references to be more frequently overspecified than singular references is related to attribute preference. As described in section 2.3, studies by (among others) Pechmann (1989) and Belke and Meyer (2002) show that speakers tend to include preferred attributes (such as 'colour') in references to one singular target, irrespective of whether these attributes have distinguishing value or not. This suggests that having more than one target may compound these effects.

2.5.2. Properties of the communicative setting

We also hypothesize that the properties of the communicative setting in which references are produced may cause referential overspecification. In this respect, we have two main hypotheses.

First, we hypothesize that spoken references are more frequently overspecified than written ones, thus containing more redundant attributes. This expectation is in line with the findings of Cohen (1984), who found that speakers provide more identification information than writers in instructive discourse. We expect similar results for the production of referring expressions, and the reason behind this is twofold. First, writing generally goes in the direction of an addressee that is not physically present, while speaking often involves a real, physically present addressee. Spoken expressions may therefore be more frequently overspecified, which would be in line with a study by Van der Wege (2009), who found instructions to an imaginary addressee to be more attenuated than instructions to a real addressee. Second, speaking, unlike writing, is an incremental process. This means that a speaker may start formulating a reference before he has fully planned what to say. We expect the same to hold in the case of spoken referring expressions, and that speakers start formulating a referring expression before having completed their scanning of the whole domain. Indeed, the incremental nature of speech production has been used to explain why language users often produce overspecified referring expressions (Pechmann, 1989). On the other hand, writers arguably are less likely to produce their expressions incrementally: they are more likely to plan the content of their references before starting to produce them, in part because a writer is typically under less time pressure than a speaker in a real-time conversation. Moreover, writers have the ability to edit the expressions they have produced, which allows them to correct their expressions for redundant attributes.

Our second hypothesis with respect to communicative setting is that speakers provide more redundant information when they cannot see the addressee than when they can. When speaker and addressee can see each other, the speaker is able to receive both auditory and visual feedback from the addressee, meaning that he can rely on the addressee's flagging the need for further information should this arise. This would fit well with the principle of mutual responsibility proposed by Clark and Wilkes-Gibbs (1986), which says that language partners collaborate in order to establish the mutual belief that the listener has understood the speaker's reference. Clark and Wilkes-Gibbs based this principle on experimental findings showing that speakers' later references to an addressee tend to contain more definite descriptions (such as "the ice skater") than early references, and also that later references contain fewer words and required fewer turns per target than early references. Clark and Wilkes-Gibbs explain these findings by claiming that speaker and addressee collaborate by establishing conceptual pacts. Arguably, this collaboration becomes more difficult when language partners cannot see each other, simply because the speaker cannot receive visual feedback from the addressee. In the absence of such feedback, a speaker may be unsure whether his or her referring expression is sufficiently precise for the addressee to single out the
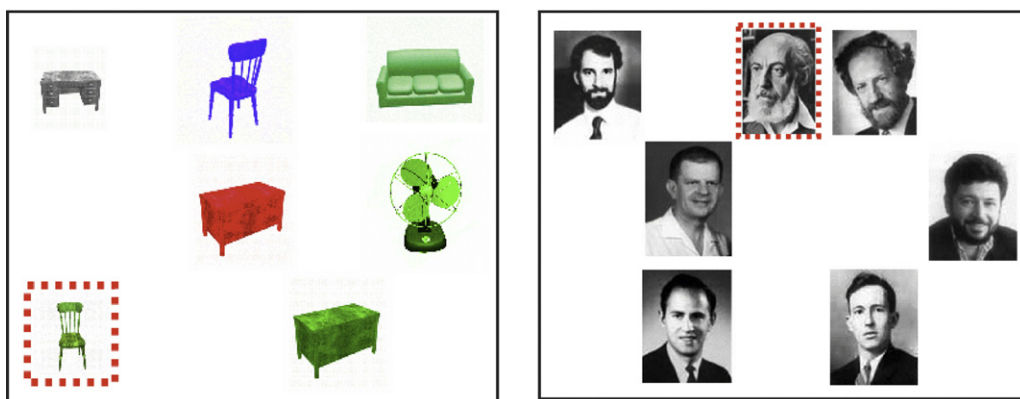
Fig. 2. Examples of (singular) trials in the furniture domain and the people domain.

target, and hence may be more inclined to add extra information. In this context, Eriksson's (2009) discussion of referential communication as an interplay between verbal and bodily practices is highly relevant. This implies that communication is less successful when the speaker and addressee cannot see each other. Further support comes from a study by Mol et al. (2009), who showed positive effects of mutual visibility on a speaker's gesture production, and, as a consequence, on the amount of information provided.

3. Which factors cause speakers to overspecify?

In the current section, we investigate to what extent properties of the target referent and properties of the communicative setting influence the amount of information included in referring expressions.

3.1. Method

The data for this study comes from a large-scale elicitation experiment with Dutch speakers, which yielded a corpus of referring expressions. The design and methodology was based on the TUNA corpus of referring expressions (Gatt et al., 2007), which was collected through an elicitation study with English speakers, who were asked to refer to targets in visual domains consisting of objects arrayed on a screen. However, while the English TUNA corpus consists entirely of written descriptions, the corpus described here, D(utch)-TUNA, contains both spoken and written expressions and, in addition, manipulates a number of other factors related to communicative setting, which we describe below.

3.1.1. Materials
The materials consisted of 40 trials, all of which contained one or two target referents and six distractor objects in a visual domain. The target referents were clearly marked by red borders, so that they could be easily distinguished from the distractor objects. For each participant and each trial, the target and distractor objects were positioned randomly on a screen in a partially filled 3 (row) by 5 (column) grid. In constructing the trials, two principal factors related to the properties of the target referent were manipulated, namely the type of domain and cardinality.

3.1.1.1. Two types of domains. A first manipulation of the target properties was that trials occurred in two different types of domains: the furniture domain and the people domain. For examples of trials in these two domains, see Fig. 2.

In each case, the attributes or dimensions along which the objects were included in the trials were defined in advance, based on previous work on the construction of the English TUNA corpus (see also Gatt et al., 2007). The 20 trials in the *furniture* domain contained artificially constructed pictures of four types of furniture items.[1] These items differed along four dimensions (see Table 1).

The 20 trials in the *people* domain consisted of pictures of males, which had been used before in an earlier study by Van der Sluis and Krahmer (2007). There were several clear differences between the two domains.[2] Most importantly, since the pictures of people were real photographs, they were not as controlled as the artificial pictures in the furniture domain. Hence, there may be more information in them that participants could use in their references, making the set of possible descriptions of a target in this domain somewhat open-ended (in that there are many unpredictable attributes that can be mentioned). However, the earlier study by Van der Sluis and Krahmer showed that speakers tend to make use of a subset of

---

[1] The pictures were taken from the Object Databank, developed by Michael Tarr at Carnegie Mellon University and freely distributed. URL: http://www.tarrlab.org/.

[2] We are aware of the fact that these motivations are specific to our stimulus material, and that they cannot be generally applied to all possible kinds of pictures of furniture items and people. We will further elaborate on this in section 4.1.1.

Table 1
Attributes and values of the furniture items.

Attribute	Possible values
Type	Chair, sofa, desk, fan
Colour	Blue, red, green, grey
Orientation	Front, back, left, right
Size	Large, small

Table 2
Attributes and values of the people pictures.

Attribute	Possible values
Type	Person
Orientation	Front, left, right
Age	Young, old
Hair colour	Dark, light
Has hair	0 (false), 1 (true)
Has beard	0, 1
Has glasses	0, 1
Has shirt	0, 1
Has tie	0, 1
Has Suit	0, 1

the available attributes much more frequently than others. Our estimate of the number of available attributes is based on this subset. Second, the target objects in the people domain could not be distinguished in terms of their type (since they are all male persons). Last, the pictures of the persons are arguably more perceptually similar to each other than the furniture items, which might make them more difficult to distinguish from the distractor objects.

As in the furniture domain, a number of salient dimensions of variation were identified in the people domain, based on previous work by Van der Sluis and Krahmer (2007) and Van der Sluis et al. (2007). These dimensions are shown in Table 2.

As in the original TUNA experiment, the construction of trials was such that, for each possible combination of the attributes in Tables 1 and 2, there was one trial in the relevant domain where that combination was minimally required in order to distinguish the referent. For example, there was a furniture trial in which the target could be distinguished using 'colour' only, another one in which the target could be distinguished using a combination of 'colour' and 'orientation', and so on. Since speakers always need to include a head noun in their references and therefore tend to always use 'type' in their formulation (Levelt, 1989), trials were built in such a way that the attribute 'type' could never be a distinguishing attribute (for example, there was no furniture trial in which a target could be uniquely distinguished by the description "The chair"). However, an attempt was made to balance the proportion of trials in the furniture domain in which objects of different types (e.g., 'chair' and 'sofa') were used as target referents.

3.1.1.2. Two levels of cardinality. A second manipulation of target properties was that trials differed in terms of cardinality, i.e. the number of target referents that they contained. Twenty trials were singular (10 per domain), containing one target referent; the remaining 20 (again 10 per domain) were plural trials containing two target referents. Fig. 2 shows examples of singular trials in the two domains, while Fig. 3 shows examples of plural trials.

Fig. 3. Examples of plural trials in the furniture domain and the people domain.
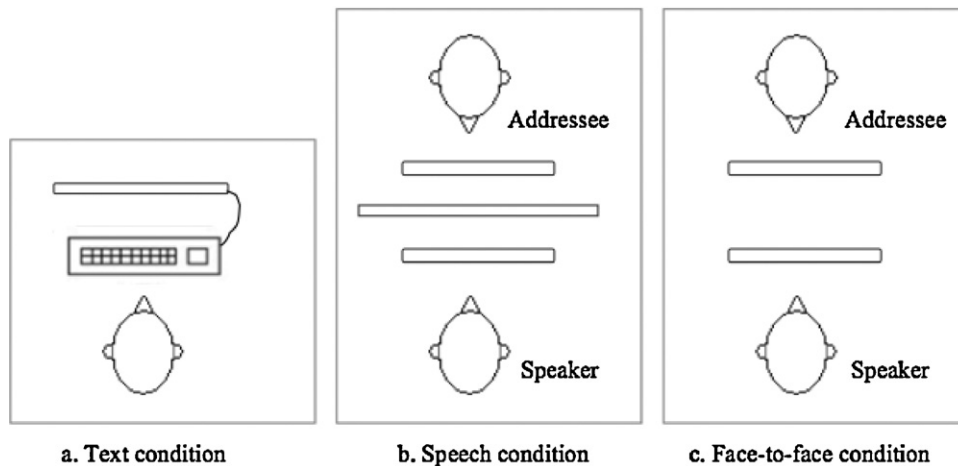
Fig. 4. Schematic overview of the three experimental conditions.

These kinds of plural trials will generally result in participants producing semantic plurals (Schwarzschild, 1996): an example of a description could be "The brown chair and the grey desk". Since plural references can be produced in different ways, we included two kinds of plural trials. Half of the trials (five per domain) contained two target objects with identical values for their distinguishing attributes. For example, both targets might be red, and 'colour' was the distinguishing attribute, so that a description like "The table and the sofa that are both red" could distinguish them. The other half (again five per domain) contained two target objects with different values on their distinguishing attributes. For example, they could consist of a fan and a sofa, both of which needed to be distinguished via their 'size' attribute, but were of different sizes. An example of a distinguishing description of such a trial would be "The large fan and the small sofa".

3.1.2. Participants

Sixty undergraduate students (14 males, 46 females) from Tilburg University participated in the experiment. All participants (mean age 20.6 years old, range 18–27 years old) were native speakers of Dutch.

3.1.3. Procedure

Each participant was presented with the same 40 trials in a different randomized order. The experiments were individually performed in an experimental room, with an average running time of 20 min. The participants could take as much time as they needed to describe the pictures. All participants were filmed during the experiment, mainly in order to capture their speech.

The participants were asked to describe the target referents in such a way that their addressee could uniquely identify them. In order to manipulate properties of the communicative setting, the participants were randomly assigned to one of three conditions (text, speech and face-to-face). The *text* condition was a replication (in Dutch) of the TUNA experiment: participants produced written identifying descriptions of the target referents. No addressee was present, but the participants were told that their descriptions would be sent to an addressee outside the experimental room. In the *speech* condition and the *face-to-face* condition, participants were asked to utter their descriptions to an addressee inside the experimental room. The addressee was a confederate of the experimenter, instructed to act as though he understood the references, but never to ask clarification questions. This was done to enable a focus on content planning of initial descriptions ('first mentions'), and to make descriptions comparable across conditions. In the instructions, the participants were told that the location of the objects on the addressee's screen had been scrambled; hence, they could not use location information in their descriptions.[3] In the face-to-face condition, the addressee was visible to the participants; in the speech condition this was not the case, because a screen was placed in between speaker and addressee. A schematic overview of the three conditions is displayed in Fig. 4.

The procedure just described differs from that used in the original TUNA experiment in three ways. First, the present study used a laboratory-based setup, whereas the original TUNA study was conducted online in a relatively uncontrolled setting. Second, as noted above, the original study only had the text condition, and participants were told that they would be writing descriptions that would be resolved in real time by an automatic language understanding system, rather than by an addressee who was in another room. Third, one of the between-groups factors manipulated in the English-language version was whether or not participants could refer to objects using their location on the screen, whereas locative descriptions were completely excluded from D-TUNA.

---

[3] See Appendix A for the verbatim instructions presented to participants in the three respective conditions.

Table 3
Overview of the experimental design and the number of descriptions within each cell.

	Furniture		People	
	Singular	Plural	Singular	Plural
Text	200	200	200	200
Speech	200	200	200	200
Face-to-face	200	200	200	200

Table 4
Examples of underspecified, minimally specified and overspecified references, with their number of words, their number of attributes (minus 'type'), and their number of redundant attributes.

Level of overspecification	Example	Total number of attributes	Number of redundant attributes
Underspecified	"The man with the beard."	1	−1
Minimally specified	"The man with the white beard."	2	0
Minimally specified	"The white bearded man."	2	0
Overspecified	"The white bearded man without a tie."	3	1

3.1.4. Data annotation

The 2400 (60 participants × 40 trials) identifying descriptions of the D-TUNA corpus were all semantically annotated in XML format, and paired with a full representation of the objects in a trial, consisting of the attributes of both the target and distractor objects. We used the annotation scheme of the TUNA corpus (Gatt et al., 2008), and carried out the annotation with the annotation tool Callisto.[4] An example of an XML representation of the people trial shown in Fig. 2 is depicted in Fig. 5, together with one of the references to this target in the corpus. In this expression, the target is referred to (in Dutch) as "*De man met de witte baard en bril*" (meaning "The man with the white beard and glasses").

All 2400 files in the corpus consist of a TRIAL node, containing a trial ID and detailing the specific conditions under which the expression was produced (such as domain, modality and cardinality). Furthermore, each TRIAL node subsumes four nodes: a DOMAIN node, a STRING-DESCRIPTION node, a DESCRIPTION node and an ATTRIBUTE-SET node.

The DOMAIN node contains a representation of the domain of the particular trial and consists of seven or eight ENTITY nodes: one or two target entities (depending on cardinality) and six distractor entities. Each entity node lists the properties of the particular entity. The STRING-DESCRIPTION node contains the full target description, as produced by the participant. The DESCRIPTION node contains the annotated version of the target description. All determiners and content words that are part of the string description were marked up with the attributes that they represent. For example, the adjective "*witte*" (meaning "white") corresponds to the attribute <hair colour = light>. In case a participant mentioned an attribute that was not predefined for the domain at all (e.g., "The laughing man"), that attribute was annotated as <other = other>. The ATTRIBUTE-SET contains an overview of all the properties (attributes and their values) that are mentioned in the string description and thus represents the "flat" semantic structure of the referring expression.

Annotating the descriptions of the D-TUNA corpus in the XML annotation scheme of the English TUNA corpus is advantageous for several reasons. First, it makes the English and the Dutch corpus highly comparable (see also section 4). Second, it makes the D-TUNA corpus a useful tool for the evaluation of algorithms that automatically generate referring expressions (REG algorithms), since the annotation is machine-readable and provides semantic information. Third, our annotation facilitates a corpus-based analysis of referential overspecification, since it explicitly marks up the semantic/conceptual content of descriptions, abstracting away from individual variation in the way these are realised syntactically.

3.1.5. Design and statistical analysis

The experiment had a 2 × 2 × 3 design (see Table 3), with two within-subjects factors: *domain* (levels: furniture, people) and *cardinality* (levels: singular, plural), and one between-subjects factor representing communicative setting: *condition* (levels: text, speech, face-to-face).

We regard the *number of redundant attributes* as a dependent variable indicating the amount of overspecification in referring expressions. An attribute is considered to be redundant if removing it from the description would still result in a distinguishing reference. For example, consider the people trial in Fig. 2 once again. Several possible descriptions of the target referent (either underspecified, minimally specified or overspecified ones) are depicted in Table 4, along with their corresponding number of redundant attributes. Since trials were built in such a way that 'type' could never be a distinguishing attribute, we excluded 'type' from our analysis of attributes.

The first reference given in Table 4, "The man with the beard", is *underspecified*. It contains only one attribute (<has beard = 1>), which is not enough for identification of the target. The next two references in Table 4, "The man with the white beard" and "The white bearded man", are *minimally specified*. They express two attributes (<has beard = 1> and <hair

---

```
<TRIAL ID="A03t21" CARDINALITY="1" CONDITION="text" DOMAIN="people" MODALITY="written">
    <DOMAIN>
        <ENTITY ID="54" IMAGE="Eilenberg.jpg" TYPE="target">
            <ATTRIBUTE NAME="hasBeard" TYPE="boolean" VALUE="1"/>
            <ATTRIBUTE NAME="hasTie" TYPE="boolean" VALUE="0"/>
            <ATTRIBUTE NAME="type" TYPE="literal" VALUE="person"/>
            <ATTRIBUTE NAME="hasHair" TYPE="boolean" VALUE="0"/>
            <ATTRIBUTE NAME="hasGlasses" TYPE="boolean" VALUE="0"/>
            <ATTRIBUTE NAME="hasSuit" TYPE="boolean" VALUE="0"/>
            <ATTRIBUTE NAME="age" TYPE="literal" VALUE="old"/>
            <ATTRIBUTE NAME="hairColour" TYPE="literal" VALUE="light"/>
            <ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="left"/>
            <ATTRIBUTE NAME="hasShirt" TYPE="boolean" VALUE="1"/>
        </ENTITY>
            <ENTITY ID="4" IMAGE="Fefferman.jpg" TYPE="distractor">
            <ATTRIBUTE NAME="hasBeard" TYPE="boolean" VALUE="1"/>
            <ATTRIBUTE NAME="hasTie" TYPE="boolean" VALUE="1"/>
            <ATTRIBUTE NAME="type" TYPE="literal" VALUE="person"/>
            <ATTRIBUTE NAME="hasHair" TYPE="boolean" VALUE="1"/>
            <ATTRIBUTE NAME="hasSuit" TYPE="boolean" VALUE="0"/>
            <ATTRIBUTE NAME="hasGlasses" TYPE="boolean" VALUE="0"/>
            <ATTRIBUTE NAME="age" TYPE="literal" VALUE="young"/>
            <ATTRIBUTE NAME="hairColour" TYPE="literal" VALUE="dark"/>
            <ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="front"/>
            <ATTRIBUTE NAME="hasShirt" TYPE="boolean" VALUE="1"/>
        </ENTITY>
            <ENTITY ID="48" IMAGE="Wall.jpg" TYPE="distractor">
            <ATTRIBUTE NAME="hasBeard" TYPE="boolean" VALUE="1"/>
            <ATTRIBUTE NAME="hasTie" TYPE="boolean" VALUE="1"/>
            <ATTRIBUTE NAME="type" TYPE="literal" VALUE="person"/>
            <ATTRIBUTE NAME="hasHair" TYPE="boolean" VALUE="1"/>
            <ATTRIBUTE NAME="hasSuit" TYPE="boolean" VALUE="1"/>
            <ATTRIBUTE NAME="hasGlasses" TYPE="boolean" VALUE="0"/>
            <ATTRIBUTE NAME="age" TYPE="literal" VALUE="young"/>
            <ATTRIBUTE NAME="hairColour" TYPE="literal" VALUE="dark"/>
            <ATTRIBUTE NAME="orientation" TYPE="literal" VALUE="front"/>
            <ATTRIBUTE NAME="hasShirt" TYPE="boolean" VALUE="0"/>
        </ENTITY>
            . . . . .
    </DOMAIN>
    <STRING-DESCRIPTION>
        De man met een witte baard en zonder bril.
    </STRING-DESCRIPTION>
    <DESCRIPTION NUM="singular">
        <DET VALUE="definite">De</DET>
        <ATTRIBUTE ID="a1" NAME="type" VALUE="person">man</ATTRIBUTE>
            met
        <ATTRIBUTE ID="a3" NAME="hasBeard" VALUE="1">een<ATTRIBUTE ID="a2"
          NAME="hairColour" VALUE="light">witte</ATTRIBUTE>baard</ATTRIBUTE>
            en
        <ATTRIBUTE ID="a4" NAME="hasGlasses" VALUE="0">bril</ATTRIBUTE>
    </DESCRIPTION>
    <ATTRIBUTE-SET>
        <ATTRIBUTE ID="a1" NAME="type" VALUE="person"></ATTRIBUTE>
        <ATTRIBUTE ID="a2" NAME="hairColour" VALUE="light"></ATTRIBUTE>
        <ATTRIBUTE ID="a3" NAME="hasBeard" VALUE="1"></ATTRIBUTE>
        <ATTRIBUTE ID="a4" NAME="hasGlasses" VALUE="0"></ATTRIBUTE>
    </ATTRIBUTE-SET>
</TRIAL>
```

Fig. 5. Example of an XML file of a reference in the people domain.

colour = light>) that are both needed for identification of the target. Therefore, no redundant attributes are counted. Finally, the fourth reference in Table 4, "The white-bearded man without a tie" is *overspecified*, since it contains one redundant attribute that is not needed for identification of the target (<has tie = 0>). Note that, as this last example illustrates, attributes can sometimes specify what is absent in the picture.

Table 5
The number of redundant attributes (means and standard deviations) as a function of all the conditions analyzed.

	Domain		Cardinality		Communicative setting		
	Furniture	People	Singular	Plural	Text	Speech	Face-to-face
Redundant attributes	.6 (.03)	1.6 (.15)	.9 (.07)	1.3 (.10)	.9 (.14)	1.3 (.14)	1.1 (.14)

Our statistical procedure consisted of mixed-model repeated measures ANOVAs, and Tukey HSD tests for post-hoc multiple comparisons. We report on main effects and interactions where these are significant.

3.2. Results

The proportions of minimally specified, overspecified, and underspecified descriptions confirmed the finding of various psycholinguistic studies that speakers tend to overspecify their referring expressions: 53.6% of the references were overspecified, as compared to 41.4% minimally specified references. Only a small minority of the references (5.0%) was underspecified. Given the fact that referential underspecification was rare in the sample (especially compared to the rates of overspecification) and did not result in significant differences on any of the relevant factors, we decided not to further address this topic in this paper.

More detailed analyses of the data indicated at least some factors that were responsible for the occurrence of referential overspecification: properties of the target referent and properties of the communicative setting. Table 5 depicts the descriptive statistics (means and standard deviations) of all the conditions analyzed.

3.2.1. Results for properties of the target

3.2.1.1. Domain. First, the results show that domain complexity can be regarded as a factor causing referential overspecification. Fig. 6 depicts the average number of redundant attributes as a function of domain.

Fig. 6 shows that references to people did contain significantly more redundant attributes ($M = 1.6$, $SD = .15$) than references to furniture items ($M = .6$, $SD = .03$), $F_{(1,57)} = 9.419$, $p < .001$. These results suggest that speakers are more likely to overspecify their references when they refer to targets that occur in a domain where the objects differ along a relatively high number of dimensions.

3.2.1.2. Cardinality. A second factor that was hypothesized to influence referential overspecification is cardinality. We first checked whether the two different kinds of plural trials differed in terms of referential overspecification, but no significant differences were found.

Fig. 7 shows the average number of redundant attributes as a function of cardinality.

Our hypothesis that references to two target objects should contain more redundant attributes than references to one target was confirmed by the significant difference between the number of redundant attributes that were mentioned in singular references ($M = .9$, $SD = .07$) and plural references ($M = 1.3$, $SD = .10$), $F_{(1,57)} = 42.714$, $p < .001$. These results indicate that speakers overspecify their references more frequently when they refer to plural targets.

The effect of cardinality described above was stronger in the more complex people domain, as reflected in interactions between domain and cardinality for the number of redundant attributes. That is, compared to the furniture domain, the effect of cardinality on the number of redundant attributes that the references contained was stronger in the people domain ($F_{(1,57)} = 30.612$, $p < .001$). In order to study this interaction in more detail, we conducted two separate statistical analyses to study whether the effect of cardinality was significant in both domains. In the furniture domain, we found an effect of cardinality ($F_{(1,57)} = 4.266$, $p < .05$), meaning that descriptions of two furniture items ($M = .64$, $SD = .05$) contained significantly more redundant target attributes than descriptions of one furniture item ($M = .55$, $SD = .03$). We found a similar (but somewhat stronger) effect of cardinality in the people domain ($F_{(1,57)} = 42.975$, $p < .001$), showing that descriptions of
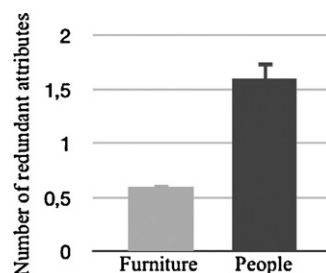

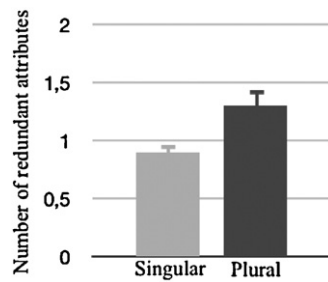**Fig. 6.** Average number of redundant attributes as a function of domain.

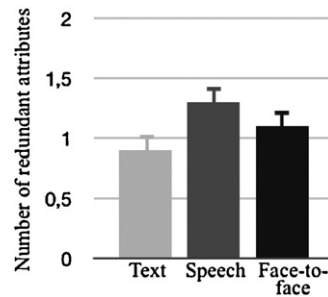**Fig. 7.** Average number of redundant attributes as a function of cardinality.

Fig. 8. Average number of redundant attributes as a function of communicative setting.

two persons ($M = 2.0$, $SD = .19$) contained significantly more redundant target attributes than target descriptions of one person ($M = 1.2$, $SD = .12$).

3.2.2. Results for properties of the communicative setting

Besides effects of properties of the target, we also looked at the influence of communicative setting on referential overspecification. Fig. 8 displays the average number of redundant attributes as a function of the three communicative conditions.

Although Fig. 8 shows differences in the average number of redundant attributes that were mentioned in the three respective conditions, suggesting that speakers in the speech condition ($M = 1.3$, $SD = .14$) were more likely to include redundant attributes in their descriptions compared to speakers in the face-to-face condition ($M = 1.1$, $SD = .14$) and to writers in the text condition ($M = .9$, $SD = .14$), these differences did not reach significance ($F_{(2,57)} = 2.668$, ns).

3.3. Summary

In this section, we have investigated whether properties of the target and properties of the communicative setting influence speakers in terms of the amount of redundant target attributes that they include in their referring expressions.

Properties of the target serve as a factor causing referential overspecification. First, we found an effect of domain, which suggests that references to targets that occur in domains in which objects afford more referential possibilities (due to there being more attributes available) contain significantly more redundant attributes than references to targets that occur in domains where speakers have fewer referential possibilities. Second, we found effects of cardinality, which indicate that references to plural targets contain significantly more redundant attributes than references to singular targets. These effects of cardinality were present in both the furniture and the people domain.

The results also show that written and spoken referring expressions do not differ in terms of overspecification: speakers do not include significantly more redundant attributes in their descriptions than people who type their expressions. The results also show that people who cannot see the addressee do not use lengthier descriptions (in terms of the number of redundant attributes) than people who can, which means that a lack of visual feedback does not directly cause speakers to overspecify their referring expressions in this type of communicative context.

4. General discussion

What makes speakers include redundant information in their referring expressions? In the current paper, we have presented a quantitative study on factors that may influence referential overspecification. The results have revealed several findings. In general, we found around 50% of the speakers' referring expressions to contain more information than needed for

identification of the target, which is in line with previous studies on referential overspecification (e.g., Engelhardt et al., 2006; Maes et al., 2004; Pechmann, 1989).

4.1. Factors causing referential overspecification

We first described a language production experiment in which we collected the D-TUNA corpus, which consists of 2400 Dutch referring expressions. These were all definite noun phrases, describing furniture items and people. We used this corpus to investigate to what extent properties of the target and properties of the communicative setting cause such definite references to be overspecificied.

4.1.1. Properties of the target referent

In line with our hypotheses, we found that properties of the target influence referential overspecification.

4.1.1.1. Domain. We have found that references to furniture items were less likely to be overspecified than references to people. These results confirm our hypothesis that the number of choices available to a speaker in the domain in which a target referent occurs affects the amount of overspecification, and that different domains can thus lead to different specification levels. This finding has at least two implications.

First, the effect of domain on referential overspecification suggests that when the range of attributes to describe a target referent gets broader (as was the case with the target objects in the people domain), speakers include more information in their target descriptions and provide the addressee with more information to single out the target. The precise effect of the number of choices available to a speaker could be studied more accurately by manipulating the number of possible attributes that is available for identifying descriptions within a single domain. This has been recently done by Koolen et al. (2011), and the results show that speakers are indeed more likely to overspecify when their referential possibilities increase.

In section 3.1.1, we have described several differences between the furniture and the people domain. One important difference was that the pictures in the people domain are more perceptually similar, which leads us to a second implication of the effect of domain on overspecification: it implies that perceptual similarity is a potentially strong indicator for the complexity of a given domain. This would mean that the complexity of a domain and the ease with which certain variables can be accessed are closely related. More specifically, it suggests that as the difficulty of selecting the discriminatory attributes increases, the speaker tends to mention more (possibly redundant) attributes in order to be sure that the addressee can identify the target referent.

It needs to be emphasized that the motivations for why we consider the furniture domain and the people domain to be different are specific to our stimulus material: they cannot be generally applied to all possible kinds of pictures of furniture items and people. For example, the situation can be different when a speaker wants to refer to people that he or she is familiar with (such as family members of friends, or Hollywood stars), or to furniture items that are all similar-looking.

4.1.1.2. Cardinality. Regarding cardinality, we would like to emphasize that the focus in this paper is on semantic plurality (Schwarzschild, 1996): rather than morphological plurals (e.g., "The brown chairs"), participants generally produced semantic plurals (e.g., "The brown chair and the grey desk"). In line with our hypothesis, we have found that plural target descriptions are more frequently overspecified than singular target descriptions. In section 2.5, we formulated two arguments that may explain the effect of cardinality on overspecification.

First, there may be competition between the two target objects, which forces speakers to divide attention. This may make the referring task more difficult, since dividing attention lowers the activation level that each of the two targets has in the speaker's mind. Such lower activation levels may cause speakers to have difficulties in avoiding redundancy, because redundancy avoidance requires comparison between a target and its distractors, and the target has to be in the focus of attention in order for this to take place successfully. While our results show that speakers often do not carry out this process of comparison exhaustively for singular targets (as shown by the high proportion of overspecified singular descriptions), these observations would also explain why this tendency appears to be even greater for plurals. This is in line with the work of Arnold and Griffin (2007), who demonstrated that plural references are more likely to contain specific forms (such as descriptions) instead of pronouns, and therefore are less attenuated than singular references. Our findings extend this observation further by highlighting another sense in which plural references are more specific than singular ones, namely in terms of the number of (redundant) attributes they contain.

Second, when referring, speakers prefer certain attributes to others. As we have observed in section 2.3, several psycholinguistic studies use attribute preference as an argument for the occurrence of referential overspecification (e.g., Pechmann, 1989; Belke and Meyer, 2002). These studies showed that speakers tend to include preferred attributes (such as 'colour') when referring to one singular target, irrespective of whether these attributes have contrastive value or not. Our findings suggest that this tendency to include preferred attributes is even greater when speakers refer to two target referents. This suggestion is in line with Gatt and van Deemter's work (2007), which shows that speakers tend to conceptualize two target objects in a parallel fashion, that is, if they describe one of two target referent using 'colour', they are bound to do the same for the other one. In the end, this leads to overspecification. For example, the plural reference "The red desk facing left and the green chair facing right" contains 'colour' twice as a preferred but redundant attribute, and is therefore more redundant than the singular reference "The red desk facing left" (which contains 'colour' as a preferred but redundant attribute only once).

4.1.2. Properties of the communicative setting

The results fail to confirm our hypotheses on the influence of the communicative setting on the redundancy level of referring expressions.

Although we have signalled a numeric trend that spoken referring expressions contain more redundant attributes than written ones, there are no significant differences between the written and the spoken conditions in terms of redundancy. This finding contrasts with previous research. Based on Cohen (1984), who found speakers to provide more identification information than writers in instruction-giving discourse, we expected this to also count for referring expressions, in the sense that written and spoken references would differ in terms of their number of redundant attributes. We further based our expectations on Van der Wege (2009), who found instructions to an imaginary addressee to be more attenuated than instructions to a real addressee. However, our results fail to confirm these expectations. One possible explanation may be that in the two spoken conditions of our experiment, the communication between speaker and addressee was rather one-sided and thus not very interactive: It was the task of the addressee to only react briefly when he had singled out the target. This lack of interaction may have reduced the role of the addressee in the two spoken conditions, in the end making it comparable to the (marginal) role of the imaginary addressee in the written condition. Some support for this interpretation comes from studies by Jordan (2000, 2002), who found that interlocutors in a collaborative dialogue task are likely to produce redundant references (including attributes which are repeated across turns, even though the referent has been established in the joint focus of attention), in order to achieve other goals in addition to identification. For example, a speaker may redundantly use the attribute 'colour' in "The red sofa" in response to her interlocutor's suggestion to "buy the red sofa", in order to signal agreement with her interlocutor about the proposal.

Even though speakers do not use more redundant attributes than writers (contrary to our expectations), they do use more words. More precisely, speakers in the speech condition ($M = 15.6$) and face-to-face condition ($M = 15.7$) used almost twice as many words than people in the text condition ($M = 8.7$). We believe that the physical presence of an addressee, as well as the incremental nature of speech production indeed causes this difference between speaking and writing. However, this did not result in mentioning more redundant attributes, which suggests that the number of words and the numbers of redundant attributes are not directly linked. An explanation for this is that speakers, more than writers, tend to repeat attributes, which may arise from disfluencies. For example, repetition of attributes in expressions such as "The red… red… chair" obviously increases the number of words used. This explanation is confirmed by the frequencies of repeated attributes in references produced in the three communicative settings. These show that spoken references in the speech and face-to-face condition contained 205 and 214 repeated attributes respectively, while written references in the text condition contained only 20 repeated attributes.

Although we have also signalled a numeric trend that speakers who cannot see an addressee provide more information than speakers who can, the referring expressions uttered in the two respective conditions did not differ in terms of the number of redundant attributes that were mentioned. This result implies that a lack of visual feedback does not lead to more referential overspecification, which contrasts with our hypothesis and with previous research (e.g., with the principle of mutual visibility by Clark and Wilkes-Gibbs, 1986). We believe that the lack of direct interaction between speaker and addressee may also explain this apparent contradiction, since it may have reduced the role of the addressee in both conditions. In future research, we aim to investigate whether this is indeed the case.

Given the lack of significant differences that we have found regarding the three communicative conditions, some concern with respect to the naturalness of our experimental setting needs to be expressed. As we have explained above, it seems that the participants in our experiment seemed not so much focused on their addressee, but more on their primary task: they had to distinguish the target referent from the contrast set of distractors. However, in naturalistic dialogue, as Jordan and Walker (2005) argue, speakers are not solely guided by this contrast set when selecting the content of their target descriptions; also other factors might play a role. First, Jordan and Walker stress the importance of intentional factors, where they explain that the attributes included in a description might have other communicative purposes than target identification alone. Furthermore, Jordan and Walker emphasize the role of the addressee in conversation, where they argue that also conceptual pact factors are important in a speaker's attribute selection process (they base themselves on papers by Brennan and Clark, 1996 and Clark and Wilkes-Gibbs, 1986 here). Since also major pragmatic theories take the role of the addressee in conversation as a starting point, we will discuss the implications of our results for these theories in the next section.

4.2. Implications of referential overspecification for pragmatic theory

One strict interpretation of the Maxim of Quantity (Grice, 1975) is that a speaker should provide just enough information for the addressee to identify a target referent (Quantity 1), but not more (Quantity 2), and that speakers should ideally produce minimally distinguishing target descriptions. However, in section 2.2, we have suggested that considering overspecification as a violation of the Maxim of Quantity would be an over-simplification: the information level of expressions generally depends on the purpose that a speaker has in uttering it. Still, several pragmatic theories on reference (which we will discuss below) argue that overspecification can have a negative effect on how an addressee comprehends an expression.

As we have seen in section 2.3, one can draw a distinction between speaker-driven and addressee-oriented factors that may cause speakers to overspecify (Arnold, 2008). Although the Maxim of Quantity intuitively seems to focus on what is said and meant by the speaker in a conversation, Grice himself makes explicit that the maxims apply to both the speaker and the addressee (Grice, 1989, p. 31). This perspective is also taken by neo-Griceans such as Horn and Levinson. Horn (1984, 2005)

mirrors the two components of the Maxim of Quantity by introducing the Q-Principle (that is addressee-oriented) and the R-Principle (that is speaker-oriented). Levinson (2000) also stresses the role of the addressee by stating that violations of the Maxim of Quantity (i.e. in case of overspecification) should be viewed as inferences drawn by the addressee. This suggests that providing redundant information could negatively affect the addressee's comprehension process.

Also Relevance theory (Carston, 1991; Sperber and Wilson, 1986; Wilson and Sperber, 1981) suggests that referential overspecification might have a negative effect on comprehension. According to this theory, utterance comprehension should be interpreted in a psychologically plausible way, which implies that an addressee always aims to arrive at an interpretation of an expression without spending much (needless) effort. In this view, expressions that meet this aim are considered 'optimally relevant'. Arguably, referring expressions that contain redundant information cannot be labeled as such, since they require unnecessary effort for the addressee to comprehend them.

In this paper, we have looked at various factors that may influence referential overspecification. We have seen that both domain properties (furniture vs. people) and target properties (singular vs. plural) may cause speakers to include more redundant properties in their referring expressions. It seems likely that this is primarily a speaker-driven process, as we have argued. In this study, we did not look explicitly at addressee-oriented aspects of referring, but it would be interesting to see how addressees process these overspecified references. As we have noted in section 2.3, how sensitive addressees are to overspecified information remains somewhat unclear. According to Grice and his followers, overspecified information may trigger false implicatures (see also Dale and Reiter, 1995). When hearing a reference to a table as "the brown wooden table", in a situation where there is only one table, an addressee might infer that the fact that the table is brown and wooden is somehow significant; why else would the speaker mention these attributes? At the same time, it seems perhaps unlikely that the additional overspecification in the case of references to people or to plural targets, creates more false implicatures, but we are not aware of any studies that directly test this.

Interestingly, we found no effects of communicative setting on the amount of overspecification. Whether references where typed or spoken did not result in significant differences in terms of the number of redundant attributes, nor did it matter whether the participant could see the addressee or nor. Although we have suggested that this result may be due to a lack of direct interaction between speakers and addressees (see section 4.1.2), it suggests that a speaker-oriented factor is involved here as well. If addressee-oriented factors would be dominant (which pragmatic theories such as Relevance theory seem to suggest), one would expect to find differences between the numbers of redundant attributes that the expressions contain in the three conditions.

4.3. Overspecification in Referring Expression Generation

As discussed at the beginning of the paper, the present paper was partly motivated by some open questions related to computational models of REG. In this section, we turn once again to this topic, and discuss some points of convergence and divergence between the terms of the present study and the computational literature, as well as the implications that our results may have for REG algorithms.

4.3.1. Definitions of redundancy

As we noted in section 2.4, concerns about overspecification and redundancy, and the tension between these and computational efficiency, have been at the heart of the development of REG algorithms. The definition of "redundant attribute" used in this paper – that is, an attribute that is redundant in relation to the rest of the description – has a precedent in the computational REG literature in Reiter's (1990) definition of *Local Brevity*. Reiter notes that a definition of redundancy as incorporated in Dale's (1989) *Full Brevity* algorithm results in a computational procedure that is intractable. Since this definition focuses on finding the smallest set of attributes that distinguish a target, it leads in the worst case to an exhaustive search through the space of all possible descriptions, starting with the shortest. Reiter's Local Brevity alternative avoids this by generating a distinguishing description and then checking whether it contains any attributes that could be removed without sacrificing identification. If this is not possible, the description is regarded as minimally specified.

The difference between the two can be illustrated through an example. Consider a domain in which "The bearded man with glasses" and "The man with the tie" both contain just enough information to make the target identifiable. That is, we assume that removing the attribute "has beard" from the first description would result in a non-identifying description, as would removing "has glasses". It is only in conjunction that these attributes identify the object. Similarly, we assume that "has tie" is required in the second description. In such a case, the Full Brevity algorithm would prefer the second description, since it contains only one attribute (besides 'type'), and therefore is the shortest description of the two. Therefore, under this definition, "The bearded man with glasses" is not strictly minimal, since there still exist shorter descriptions (such as "The man with the tie"). By contrast, under a Local Brevity assumption, both of these would qualify as "non-redundant" descriptions, since neither contains attributes that can be removed while satisfying the identification goal.

Clearly, there will be cases where the two definitions converge, but it is worth noting that from a computational perspective, only the Local Brevity assumption is tractable in the worst case. The definition of redundant attribute used in this paper is in line with the Local Brevity assumption, although separate analyses of the same data show that defining referential overspecification in line with Full Brevity also evinces similar trends.

The above discussion should not of course be taken to imply that either of the two algorithmic interpretations of redundancy discussed here is psycholinguistically plausible, but this does not hamper the impact of this study, which does

rely on any such algorithmic interpretation. For one thing, neither of these two algorithms is incremental, and incrementality is one of the core aspects of psycholinguistic models that seek to explain referential overspecification (e.g., Pechmann, 1989). As discussed in section 2.4, incremental REG procedures have since been proposed – notably by Dale and Reiter (1995) – and have been claimed to provide a better approximation to what people do, in part because their output may be overspecified. It is therefore worth exploring the possible implications that our findings have for such algorithms.

4.3.2. Computational implications of referential overspecification

Some psycholinguistic research suggests that referential overspecification can be beneficial to readers or listeners, since it facilitates the target identification process (e.g., Arts et al., 2011). This would suggest that automatically generated referring expressions containing redundant attributes would be easier to resolve. For example, recent work by Paraboni et al. (2006) shows that an optimal level of overspecification can help readers to resolve references to parts of documents. On the other hand, current REG algorithms, including the Incremental Algorithm of Dale and Reiter (1995), do not systematically generate overspecified expressions. For example, the Incremental Algorithm discussed in section 2.4 depends on an a priori specification of a preference ordering of attributes, but will only add an attribute (even one which is highly preferred) if it has discriminatory value, that is, if it excludes some of the remaining distractors of the target. Thus, this algorithm will not use 'colour' for a target in a domain in which all other objects have the same colour. Moreover, the algorithmic implications of such factors as domain complexity have not, to our knowledge, been addressed. How does the present study contribute to improving the performance of REG algorithms in terms of the amount of information they should include in their referring expressions?

On the basis of our empirical findings, we can formulate several important implications for automatic Referring Expression Generation. First, REG algorithms should include more redundant information when referring to more complex target referents, based on effects of domain and cardinality on referential overspecification in our human corpus data. More specifically, this means that automatically generated referring expressions to two target referents that are uttered in a complex domain should contain more redundant information than those to one target referent that are uttered in a simpler domain. Second, our findings suggest that modality (text or speech) and whether or not a speaker has the possibility to give visual feedback should not influence the redundancy level of automatically generated referring expressions (but see the caveats raised in the previous section).

How can REG algorithms determine which attributes should be redundantly added to their referring expressions? Based on Maes et al. (2004), we recommend that the kinds of properties that are used to describe a target should be of a perceptual nature. In particular, speakers tend to include perceptually salient target attributes (such as colour) in their references, because both speaker and addressee easily perceive such attributes (Belke and Meyer, 2002). Since Arts (2004) and Arts et al. (2011) found that locative expressions also facilitate object identification, presumably because they help to orient attention to the right location in space, it would seem to make sense to redundantly add location information to automatically generated referring expressions in such spatial domains. Stipulating that algorithms should include perceptually salient or spatially useful attributes even when they are redundant has some implications for algorithm design. In particular, it would seem that a strategy such as that incorporated in the Incremental Algorithm, where salient attributes are only included if they have some discriminatory value, might not always return results which are compatible with what humans would do in the same situation. On the other hand, it is plausible to assume that humans exhibit a degree of non-determinism in their referential behaviour, so that the inclusion of, say 'colour' in a domain where all objects are of the same colour, may be probabilistic.

As described in section 2.4, evaluation of REG algorithms is not only a case of aiming to emulate human referring behaviour, but also to produce *effective* referring expressions (van Deemter et al., 2010), that is, expressions which are easy for an addressee to comprehend and/or resolve (Paraboni et al., 2007). Previous research has shown that including redundant information in a referring expression facilitates object identification (e.g., Arts et al., 2011; Deutsch, 1976; Nadig and Sedivy, 2002; Engelhardt et al., 2006; Paraboni et al., 2006). This implies that REG algorithms might include redundant information in their referring expressions, because this makes them easier to comprehend for addressees. But what level of redundancy is optimal in any given situation? Arguably, pragmatic factors related to the communicative setting are likely to affect the amount of redundant information (van Deemter et al., 2007). On the one hand, there may be fault-critical settings, where accurate understanding is crucial; On the other hand, there may be situations in which the reader of listener's ease of comprehension is not at the forefront of the speaker or writer's attention, for instance when the speaker or writer is under time pressure. It stands to reason that redundancy is more important in the former case.

5. Conclusion

The current study addresses several factors that cause speakers to overspecify their referring expressions, or, more specifically, the definite target descriptions that they produce. The results of our language production experiment show that the domain in which a target occurs, and whether this target is singular or plural, determines the amount of redundant information that language users include in their definite descriptions. However, the results do not show an effect of communicative setting (concerning the modality in which an expression is produced, and the speaker's possibility to receive visual feedback from the addressee) on overspecification. This was possibly due to a lack of natural interaction between

speaker and addressee. Since current REG algorithms are not able to deal with referential overspecification in a systematic way, the above pragmatic findings are helpful in improving these algorithms in terms of the amount of information that they should include in their referring expressions.

Acknowledgements

Appendix A. Verbatim instructions to the participants (English translations of Dutch originals)

Text condition

In this experiment we collect descriptions that people give when they have to describe an object, for example: "*The blue car with an open roof*". You will be asked to describe pictures in such a way that your addressee can uniquely identify them. The procedure is as follows:

During the experiment, you will be repeatedly presented with a collection of pictures. These are pictures of **furniture items** and **people**. In each collection of pictures, one or two objects are marked by red boarders. It is your task to **describe these marked pictures in such a way that they can be distinguished from the other pictures on the screen**.

You will be filmed during the experiment, and your descriptions will be sent to a person in another room. This person has the task to identify the pictures that he or she thinks you are referring to. This person can see you typing your descriptions, but does not hear anything.

The person in the other room is presented with the same pictures as you are. However, the pictures that he or she sees are in a different position than your objects. This means that **it does not make any sense to describe the pictures in terms of their position on the screen** (e.g., "The chair in the center").

We would like to ask you to describe the objects in such a way that they sound natural to you as a native speaker of Dutch. You can type your description in the textbox beneath the objects. Every time you have finished a description, you click on 'next'. Once you have done that, the description can no longer be changed. You can take as much time as you like for each description.

Speech condition

In this experiment we collect descriptions that people give when they have to describe an object, for example: "*The blue car with an open roof*". You will be asked to describe pictures in such a way that your addressee can uniquely identify them. The procedure is as follows:

During the experiment, you will be repeatedly presented with a collection of pictures. These are pictures of **furniture items** and **people**. In each collection of pictures, one or two objects are marked by red boarders. It is your task to **describe these marked pictures in such a way that the addressee can distinguish them from the other pictures on the screen**. The addressee is presented with the same pictures as you are. However, the pictures that he or she sees are in a different position than your objects. This means that **it does not make any sense to describe the pictures in terms of their position on the screen** (e.g., "The chair in the center").

We would like to ask you to describe the objects in such a way that they sound natural to you as a native speaker of Dutch. Every time you have finished a description, you click on 'next'. Once you have done that, the description can no longer be changed. You can take as much time as you like for each description.

There is a screen in between you and your addressee, so that you cannot see each other. The addressee can however hear your descriptions.

You will be filmed during the experiment, so that we will be able to hear back your descriptions.

Face-to-face condition

Participants in the face-to-face conditions were presented with the same instructions as the ones in the speech condition, except for the fact that the sentence describing that a screen was placed in between speaker and addressee was deleted.

References

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

Arnold, Jennifer E., Griffin, Zenzi M., 2007. The effect of additional characters on choice of referring expressions: Everyone competes. Journal of Memory and Language 56, 521–536.

Arts, Anja, 2004. Overspecification in instructive texts. Dissertation, Tilburg University. Wolf Publishers, Nijmegen.

Arts, Anja, Maes, Alfons, Noordman, Leo, Jansen, Carel, 2011. Overspecification facilitates object identification. Journal of Pragmatics 43 (1), 361–374.

Bach, Kent, 1994. Thought and Reference. Oxford University Press, Oxford.

Bach, Kent, 2006. The top ten misconceptions about implicature. In: Horn, R., Birner, B., Ward, G. (Eds.), Drawing the Boundaries of Meaning: Neo-Gricean Studies in Pragmatics and Semantics in Honor of Laurence R. John Benjamins, Amsterdam, pp. 21–30.

Belke, Eva, Meyer, Antje, 2002. Tracking the time course of multidimensional stimulus discrimination: analysis of viewing patterns and processing times during "same"–"different" decisions. European Journal of Cognitive Psychology 14, 237–266.

Brennan, Susan E., Clark, Herbert H., 1996. Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: Learning, Memory and Cognition 22 (6), 1482–1493.

Byron, Donna, Koller, Alexander, Striegnitz, Kristina, Cassell, Justine, Dale, Robert, Moore, Johanna, Oberlander, Jon, 2009. Report on the first NLG Challenge on Generating Instructions in Virtual Environments (GIVE). In: Proceedings of the 12th European Workshop on Natural Language Generation, Athens, Greece.

Carston, Robyn, 1991. Implicature, explicature and truth-theoretical semantics. In: Davis, S. (Ed.), Pragmatics. Oxford University Press, New York.

Clark, Herbert H., Wilkes-Gibbs, Deanna, 1986. Referring as a collaborative process. Cognition 22, 1–39.

Cohen, Philip R., 1984. The pragmatics of referring and the modality of communication. Computational Linguistics 10, 97–146.

Dale, Robert, 1989. Cooking up referring expressions. In: Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics, University of British Columbia, Vancouver, BC, Canada, pp. 68–75.

Dale, Robert, 1992. Generating Referring Expressions: Building Descriptions in a Domain of Objects and Presses. MIT Press, Cambridge.

Dale, Robert, Reiter, Ehud, 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. Cognitive Science 18, 233–263.

Davies, Catherine, Katsos, Napoleon, 2009. Are interlocutors as sensitive to over-informativeness as they are to under-informativeness? In: Proceedings of the CogSci workshop on the Production of Referring Expressions (PRE-CogSci 2009), Amsterdam, The Netherlands.

Deutsch, Werner, 1976. Sprachliche Redundanz und Objekt Identifikation. University of Marburg, Dissertation, Marburg.

Deutsch, Werner, Pechmann, Thomas, 1982. Social interaction and the development of definite descriptions. Cognition 11, 159–184.

Eberhard, Kathleen M., Spivey-Knowlton, Michael J., Sedivy, Julie C., Tanenhaus, Michael K., 1995. Eye movements as a window into real-time spoken language comprehension in natural contexts. Journal of Psycholinguistic Research 24, 409–436.

Eikmeyer, Hans-Jürgen, Ahlsén, Elisabeth, 1996. The cognitive process of referring to an object: a comparative study of German and Swedish. In: Proceedings of the 16th Scandinavian Conference on Linguistics, Turku, Finland.

Engelhardt, Paul E., Bailey, Karl G.D., Ferreira, Fernanda, 2006. Do speakers and listeners observe the Gricean maxim of quantity? Journal of Memory and Language 54, 554–573.

Engelhardt, Paul E., Demiral, Şikrü Bariş, Ferreira, Fernanda, 2010. Over-specified referential expressions impair comprehension: an ERP study. In: Proceedings of the 16th conference on Architectures and Mechanisms for Language Processing (Amlap), York, England.

Eriksson, Mats, 2009. Referring as interaction: on the interplay between linguistic and bodily practices. Journal of Pragmatics 41 (2), 240–262.

Ford, William, Olson, David, 1975. The elaboration of the noun phrase in children's description of objects. Journal of Experimental Child Psychology 19 (3), 371–382.

Gardent, Claire, 2002. Generating minimal definite descriptions. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, USA, pp. 96–103.

Gatt, Albert, 2007. Generating coherent references to multiple entities. Unpublished PhD thesis, University of Aberdeen.

Gatt, Albert, Belz, Anja, 2010. Introducing shared task evaluation to NLG: the TUNA shared task evaluation challenges. In: Krahmer, E., Theune, M. (Eds.), Empirical methods in Natural Language Generation. Springer, Berlin and Heidelberg.

Gatt, Albert, van Deemter, Kees, 2007. Incremental generation of plural descriptions: similarity and partitioning. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2007), Prague, Czech Republic.

Gatt, Albert, van der Sluis, Ielka, van Deemter, Kees, 2007. Evaluating algorithms for the generation of referring expressions using a balanced corpus. In: Proceedings of the Eleventh European Workshop on Natural Language Generation, Saarbruecken, Germany, pp. 49–56.

Gatt, Albert, van der Sluis, Ielka, van Deemter, Kees, 2008. XML formatting guidelines for the TUNA corpus. Technical report. Department of Computing Science, University of Aberdeen.

Gatt, Albert, Portet, Francois, Reiter, Ehud, Hunter, Jim, Mahamood, Saad, Moncur, Wendy, Sripada, Somayajulu, 2009. From data to text in the neonatal intensive care unit: using NLG technology for decision support and information management. AI Communications 22, 153–186.

Goldberg, Eli, Driedger, Norbert, Kittredge, Richard I., 1994. Using Natural-Language Processing to produce weather forecasts. IEEE Expert: Intelligent Systems and Their Applications 9 (2), 45–53.

Goudbeek, Martijn, Krahmer, Emiel, 2010. Preferences versus adaptation during Referring Expression Generation. In: Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Uppsala, Sweden, pp. 55–59.

Grice, H. Paul, 1975. Logic and conversation. In: Cole, P., Morgan, J.L. (Eds.), Speech Acts. Academic Press, New York, pp. 41–58.

Grice, H. Paul, 1989. Studies in the Way of Words. Harvard University Press, Cambridge, MA.

Gupta, Surabhi, Stent, Amanda, 2005. Automatic evaluation of referring expression generation using corpora. In: Proceedings of the 1st Workshop on Using Corpora in Natural Language Generation, Brighton, UK, pp. 1–6.

Heeman, Peter, Hirst, Graeme, 1995. Collaborating on referring Expressions. Computational Linguistics 21 (3), 282–351.

Horn, Laurence R., 1984. Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In: Schriffrin, D. (Ed.), Meaning, Form and Use in Context: Linguistic Applications. Georgetown University Press, Washington.

Horn, Laurence R., 2005. Current issues in neo-Gricean pragmatics. Intercultural Pragmatics 2, 191–204.

Jordan, Pam, 2000. Influences on attribute selection in redescriptions: a corpus study. In: Proceedings of the 22nd Annual Conference of the Cognitive Science Society, Pennsylvania, USA, pp. 250–255.

Jordan, Pam, 2002. Contextual Influences on attribute selection for repeated descriptions. In: van Deemter, K., Kibble, R. (Eds.), Information Sharing: Reference and Presupposition in Natural Language Generation and Understanding. CSLI, Stanford, CA.

Jordan, Pam, Walker, Marilynn, 2005. Learning content selection rules for generating object descriptions in dialogue. Journal of Artificial Intelligence Research 24, 157–194.

Kelleher, John, Kruijff, Gert-Jan, 2006. Incremental generation of spatial referring expressions in situated dialogue. In: Proceedings of the Joint 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics.

Koolen, Ruud, Goudbeek, Martijn, Krahmer, Emiel, 2011. Scene variation affects referential overspecification. In: Proceedings of the 33rd annual meeting of the Cognitive Science Society (CogSci), Boston, USA.

Krahmer, Emiel, van Erk, Sebastiaan, Verleg, André, 2003. Graph-based Generation of Referring Expressions. Computational Linguistics 29 (1), 53–72.

Levelt, Willem J.M., 1989. Speaking: From Intention to Articulation. MIT Press, Cambridge/London.

Levinson, Stephen C., 1979. Activity types and language. Linguistics 17, 365–399.

Levinson, Stephen C., 2000. Presumptive Meaning. The Theory of Generalized Conversational Implicature. MIT Press, Cambridge MA.

Maes, Alfons, Arts, Anja, Noordman, Leo, 2004. Reference management in instructive discourse. Discourse Processes 37 (2), 117–144.

Mellish, Chris, Scott, Donia, Cahill, Lynn, Evans, Roger, Paiva, Daniel, Reape, Mike, 2006. A reference architecture for Natural Language Generation systems. Natural Language Engineering 12 (1), 1–34.

Mol, Lisette, Krahmer, Emiel, Maes, Alfons, Swerts, Marc, 2009. The communicative import of gestures: evidence from a comparative analysis of human-human and human-machine interactions. Gesture 9 (1), 98–127.

Mooney, Annabelle, 2004. Co-operation, violations and making sense. Journal of Pragmatics 36 (5), 899–920.

Nadig, Aparna S, Sedivy, Julie C., 2002. Evidence of perspective-taking constraints in children's on-line reference resolution. Psychological Science 13, 329–336.

Olson, David R., 1970. Language and thought: aspects of a cognitive theory of semantics. Psychological Review 77, 257–273.

Paraboni, Ivandré, Masthoff, Judith, van Deemter, Kees, 2006. Overspecified reference in hierarchical domains: Measuring the benefits for readers. In: Proceedings of the 4th International Conference on Natural Language Generation (INLG-06), Sydney, Australia, pp. 55–62.

Paraboni, Ivandré, van Deemter, Kees, Masthoff, Judith, 2007. Making referents easy to identify. Computational Linguistics 33 (2), 229–254.

Pechmann, Thomas, 1989. Incremental speech production and referential overspecification. Linguistics 27, 89–110.

Pickering, Martin, Garrod, Simon, 2004. Towards a mechanistic psychology of dialogue. Behavioural and Brain Sciences 27, 169–226.

Poesio, Massimo, Vieira, Renata, 1998. A corpus-based investigation of definite reference use. Computational Linguistics 24, 183–216.

Reiter, Ehud, 1990. The computational complexity of avoiding conversational implicatures. In: Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics (ACL-1990), Pittsburgh, Pennsylvania, USA, pp. 97–104.

Reiter, Ehud, Dale, Robert, 2000. Building Natural Language Generation systems. Cambridge University Press, Cambridge.

Reiter, Ehud, Sripada, Somayajulu, Hunter, Jim, Yu, Jin, Davy, Ian, 2005. Choosing words in computer-generated weather forecasts. Artificial Intelligence 167, 137–169.

Schriefers, Herbert, Pechmann, Thomas, 1988. Incremental production of referential noun phrases by human speakers. In: Zock, M., Sabah, G. (Eds.), Advances in Natural Language Generation, vol. 1. Pinter, London, pp. 172–179.

Schwarzschild, Roger, 1996. Plurality. Kluwer, Dordrecht, The Netherlands.

Searle, John R., 1969. Speech Acts: An Essay in the Philosophy of Language. Oxford University Press, Oxford.

Sedivy, Julie C., Tanenhaus, Michael K., Chambers, Craig G., Carlson, Gregory N., 1999. Achieving incremental semantic interpretation through contextual representation. Cognition 71, 109–147.

Sonnenschein, Susan, 1984. The effect of redundant communication on listeners: why different types may have different effects. Journal of Psycholinguistic Research 13, 147–166.

Sonnenschein, Susan, Whitehurst, Grover J., 1982. The effects of redundant communications on the behavior of listeners: does a picture need a thousand words? Journal of Psycholinguistic Research 11 (2), 115–125.

Sperber, Dan, Wilson, Deidre, 1986. Relevance: Communication and Cognition. Harvard University Press, Cambridge, MA.

Spivey, Michael J., Richardson, Daniel C., 2008. Language embedded in the environment. In: Robbins, P., Aydede, M. (Eds.), The Cambridge Handbook of Situated Cognition. Cambridge University Press, Cambridge, UK, pp. 383–400.

Stock, Oliviero, Zancanaro, Massimo, Busetta, Paolo, Callaway, Charles, Krüger, Antonio, Kruppa, Michael, Kuflik, Tsvi, Not, Elena, Rocchi, Cesare, 2007. Adaptive, intelligent presentation of information for the museum visitor in PEACH. User Modeling and User-Adapted interaction 17 (3), 257–304.

Stoia, Laura, Shockley, Darla Magdalene, Byron, Donna K., Fosler-Lussier, Eric, 2006. Noun phrase generation for situated dialogs. In: Proceedings of the 4th International Conference on Natural Language Generation, Morristown, NJ, USA, pp. 81–88.

Tanenhaus, Michael K., Spivey-Knowlton, Michael J., Eberhard, Kathleen M., Sedivy, Julie C., 1995. Integration of visual and linguistic information in spoken language comprehension. Science 268, 1632–1634.

Van Deemter, Kees, 2002. Generating referring expressions: Boolean extensions of the incremental algorithm. Computational Linguistics 28 (1), 37–52.

Van Deemter, Kees, Gatt, Albert, Van Gompel, Roger, Krahmer, Emiel. Towards a computational psycholinguistics of reference production. Topics in Cognitive Science, in press.

Van der Sluis, Ielka, Krahmer, Emiel, 2007. Generating multimodal references. Discourse Processes 44, 145–174.

Van der Sluis, Ielka, Gatt, Albert, Van Deemter, Kees, 2007. Evaluating algorithms for the generation of referring expressions: going beyond toy domains. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP-07), Borovets, Bulgaria.

Van der Wege, Mija M., 2009. Lexical entrainment and lexical differentiation in reference phrase choice. Journal of Memory and Language 60, 448–463.

Viethen, Jette, Dale, Robert, 2006. Algorithms for generating referring expressions: do they do what people do? In: Proceedings of the 4th International Conference on Natural Language Generation (INLG-06), Sydney, Australia, pp. 63–70.

Wilson, Deidre, Sperber, Dan, 1981. On Grice's theory of conversation. In: Werth, P. (Ed.), Conversation and Discourse. Croom Helm, London.

Yip, Virginia Choi-Yin, Matthews, Stephen, 2007. The Bilingual Child: Early Development and Language Contact. Cambridge University Press, Cambridge.

Ruud Koolen is a PhD student at the Department of Communication and Information Sciences, Faculty of Humanities at Tilburg University, the Netherlands. His research concerns getting a better understanding of human reference production and comprehension, as well as formulating implications for automatic Referring Expression Generation (REG) algorithms.

Albert Gatt is a lecturer in the Institute of Linguistics at the University of Malta, as well as a visiting researcher at the Department of Communication and Information Sciences, Faculty of Humanities at Tilburg University, the Netherlands. His research focuses on psycholinguistic and computational models of natural language production.

Martijn Goudbeek is an assistant professor at the Department of Communication and Information Sciences, Faculty of Humanities at Tilburg University, the Netherlands. His research concerns the role of entrainment and discourse processes in human production of referring expressions to inform developments in algorithms for automatic REG in dialogue.

Emiel Krahmer is a full professor at the Department of Communication and Information Sciences, Faculty of Humanities at Tilburg University, The Netherlands. His research is aimed at getting a better understanding of how humans exchange information during communication (both verbally and non-verbally, using speech, gestures and facial expressions), which in turn may help for improving the way computers present information and communicate with humans.