Journal of Pragmatics 216 (2023) 21–42

Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: www.elsevier.com/locate/pragma

Varieties of specification: Redefining over- and under-specification

Guanyi Chen*, Kees van Deemter

Department of Information and Computing Sciences, Utrecht University, Utrecht, the Netherlands

ARTICLE INFO

Article history:
Received 1 February 2023
Received in revised form 28 July 2023
Accepted 28 July 2023

Keywords:
Referring expressions
Over-specification
Under-specification

ABSTRACT

A long tradition of research in theoretical, experimental and computational pragmatics has investigated over-specification and under-specification in referring expressions. Along broadly Gricean lines, these studies compare the amount of information expressed by a referring expression against the amount of information that is required. Often, however, these studies offer no formal definition of what "required" means, and how the comparison should be performed. In this paper, we use a simple set-theoretic perspective to define some communicatively important types of over-/under-specification. We argue that our perspective enables an enhanced understanding of reference phenomena that can pay important dividends for the analysis of reference in corpora and for the evaluation of computational models of referring. To illustrate and substantiate our claims, we analyse two corpora, containing Chinese and English referring expressions respectively, using the new perspective. The results show that interesting new monolingual and cross-linguistic insights can be obtained from our perspective.

1. Introduction

The primary function of a referring expression is to help hearers identify what a speaker is talking about: the intended referent. Referring expressions have been studied from many angles, including linguistic, psychological, and computational. The aim of the present paper is to propose a new, more precise way of thinking about the extent to which, and the manner in which, a given referring expression achieves its primary function. This perspective will make finer distinctions than usual, and place well-known ideas, such as referential over-specification, within a wider spectrum of referential behaviours, including both under-specification and erroneous specification. We will argue that this enhanced perspective will offer (1) a better understanding of reference itself, (2) a more useful way to annotate referring expressions in corpora, and (3) a better way of comparing the referring expressions that may be produced in a given situation. As a by-product, it offers a more refined way of measuring the success of algorithms that seek to model (i.e., mimic) human production of referring expressions.

A long tradition of studies (Olson, 1970; Ford and Olson, 1975; Sonnenschein, 1984; Pechmann, 1989; Engelhardt et al., 2006, 2011; Koolen et al., 2011; Paraboni et al., 2017; Degen et al., 2020) of human reference production has paid attention to the human tendency to produce referring expressions that contain more semantic information than is strictly necessary for identifying the intended referent. Researchers frequently refer to such expressions as over-specified referring expressions or *over-specifications* for short (Pechmann, 1989). For instance, experiments have shown that in the situation of

* Corresponding author. Buys Ballotgebouw, Princetonplein 5, Room 5.01, Utrecht, 3584CC, the Netherlands.
E-mail addresses: g.chen@ccnu.edu.cn (G. Chen), c.j.vandeemter@uu.nl (K. van Deemter).

https://doi.org/10.1016/j.pragma.2023.07.019

Fig. 1, many speakers tend to say "the large green chair" to identify the referent, even though shorter descriptions, such as "the green chair", and "the large chair", would suffice. Expressions such as the latter two, which contain the minimum number of properties (two properties, in these two cases) required for identifying the referent, are called minimally specified, or simply minimal. Precise definitions of these notions will follow in Section 3.

Fig. 1. A scene that requires speakers producing referring expressions to single out the object in the window from others. For added clarity, the colour of each object is written below it. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

A famous attempt to explain the idea of rational communication is embodied in the *Gricean Maxims* (Grice, 1975). Particularly relevant for present purposes is the Maxim of Quantity, which is comprised of two rules:

1. the speaker should make the contribution as informative as required;
2. the speaker should not make the contribution more informative than is required.

Later researchers have defined over-specification in essentially Gricean terms (Pechmann, 1989; Engelhardt et al., 2006, 2011; Paraboni et al., 2017; Dale and Reiter, 1995; Koolen et al., 2011): for these researchers, if an expression violates the second rule of quantity, it is considered to be an over-specification. If an expression breaks the first rule, it is considered an under-specification.

This approach, which has informed many theories and computational models, and in which the expression "(as informative as) required" takes a central position, has a number of limitations: most obviously, it does not specify what or how much information actually is required. Equally importantly, the above approach lumps together situations that are intuitively very different, designating them all as over-specifications. These claims will be argued at length in the following sections.

Let us elaborate, starting with the first of the above-mentioned limitations. To refer to the target in Fig. 1, speakers say such things as:

(1) a. the large one
b. the green chair
c. the large chair
d. the large green one
e. the large blue chair in the middle.
f. the green chair that has the same colour as the desk.

Both (1-a) and (1-b) provide all the "required" information, allowing hearers to identify the target object. They contain no redundant information since none of their content words could be omitted. Nevertheless, the description (1-b) mentions two attributes (i.e., the COLOUR green and TYPE chair) whereas (1-a) mentions only one (i.e., the SIZE large), suggesting that, in an obvious sense, the former conveys more information than required. Many studies leave it unclear whether or not such cases involve over-specification.

It is even less clear what one should make of (1-e), where something patently false is said about the chair, and (1-f), where the colour of the chair is expressed twice, though in different ways; whether or not such expressions should be regarded as over-specified is often not clear from the literature. Expressions of these and kinds are common in the corpora that we will study in later sections.

Furthermore, an utterance such as (1-c) is technically an over-specification, because the property of being a chair can be removed without causing referential ambiguity. Yet a long history of experimentation (Levelt, 1993, Chapter 4) shows that descriptions like (1-c) are extremely common. This is thought to be because the TYPE of a referent — simply put, the information contributed by the noun — plays a special role, because "chair" helps the speaker to construct a grammatically correct noun phrase (NP) in English, perhaps *via* a *Gestalt* in the human mind (see Levelt (1993) for explanation). A proper analysis of over-specification should therefore distinguish between over-specifications that are caused merely by the presence of logically superfluous TYPE attributes and over-specifications that are caused by other attributes.

Another limitation of existing approaches is that they do not support quantitative analysis, which would tell us *how much* of the information in an overspecified Noun Phrase is surplus to requirements. A proper quantitative analysis could plausibly contribute to understanding the communicative style of a particular person or demography, such as children (Ford and Olson, 1975; Matthews et al., 2016) or people on the autism spectrum (e.g., Pogue et al. (2016)), which is something that previous

accounts have struggled to do. Our experiments in Section 4 show the usefulness of quantitative analysis in a more theoretical scenario, in which we ask to what extent languages such as English and Mandarin use different referential strategies.

To address all these issues, we take a fresh look at over-specification and a bunch of other phenomena that are closely related to over-specification. We develop a new perspective that is *clearer* than earlier accounts, because it uses explicit definitions for all its key terms; it is *more general* because it allows that information can not only be added but also omitted or distorted; it is also *more fine grained* because we subcategorize the notion of over-specification and we provide a precise definition for each category.

As an umbrella term for this entire family of phenomena, we use the term *specification*. Using our new perspective on specification, we annotated a corpus called MTUNA (van Deemter et al., 2017) of referring expressions in Mandarin, to see how speakers of Mandarin – where reference works differently from English (see Chen (2022)) – use over- and under-specification.

Our work stands in a research tradition whose aim is to build empirically grounded and linguistically insightful computational models of language use. Such models are often either rule-based or based on classic (i.e., interpretable) Machine Learning, avoiding the "black box" models that have dominated much recent work in Natural Language Processing. The construction of a computational model requires that linguistic concepts are defined explicitly enough to enable us to formulate and test statistical hypotheses and to implement algorithms on a computer, which can subsequently be evaluated rigorously. The requirements of using explicit definitions, and evaluating algorithms rigorously, can cause the algorithms in question to appear simple-minded in comparison to linguistic competitors, which tend to address more complicated and unusual types of language use. This is because the mindset of computational modelling is to get the simple things right before the harder things are attempted. The benefits of this approach include formal rigour and experimental support.

In this paper, we focus on relatively simple referential situations in which it is always clear whether or not an object has a certain property (e.g. whether a desk is red, whether or not it is small). Debatable and vague properties are thus left out of consideration.[1] Similarly, we will disregard the wider context of an NP, because when context is taken into account, this can often affect interpretation. (We can say "*the dog*" to refer to a particular dog, even though there are many dogs in the world, as long as the intended referent is the contextually most salient dog (Krahmer and Theune, 2002).) This will allow us to keep our definitions straightforward. In Section 5.3.1, we will discuss how our ideas apply to more complicated communicative situations in which the wider context of a referring expression cannot be ignored.

This paper is organised as follows: in Section 2, we will introduce the main concepts, methods, and findings in this area. In Section 3, we offer a formal explication of the perspective we are proposing. Section 4 puts our new perspective to work, analysing some actual corpora. Section 5 sums up our findings, discusses some limitations of our approach and draws conclusions.

2. Varieties of specification in different research traditions

The study of referring can be broadly subdivided into three different research traditions: theoretical linguistic, computational, and psycholinguistic. (For connections between these traditions, and discussion, see van Deemter (2016). We summarise each of the three, focusing on work of particular relevance to issues of over- and under-specification.

2.1. Reference in philosophy and linguistics

The idea of a referring expression was aptly summarised by John Searle, who wrote "Any expression which serves to identify anything, process, event, action, or any kind of individual or particular I shall call a referring expression. Referring expressions point to particular things; they answer the questions Who?, What?, Which?" (Searle, 1969). The study of referring expressions (see van Deemter (2016), Chapter 2, for an overview) can be traced back to John Stuart Mill's distinction between connotation and denotation (Mill, 1843) and Frege's related notions of *Sinn* and *Bedeutung* (Frege, 1892). It gained traction following Strawson's critique of Russell's theory of descriptions (Russell, 1905; Strawson, 1950).

The formal semantics literature contains a famous debate about the question of whether a sentence that contains a referring expression is best seen as asserting (Russell) or presupposing (Strawson) the uniqueness of the referent. Many of these debates are irrelevant to our present purposes. For example, whether a sentence like "*The King of France is bald*" *asserts* or *presupposes* the existence of exactly one King of France, the question of whether the description "*The King of France*" is over- or under-specified remains the same. This is also true for the bulk of work on intensionality, which has revolved around the question under what circumstances two expressions are interchangeable *salva veritate* (e.g., Neale (1988)).

An important impetus to thinking about the felicity of using a referring expression was given when Paul Grice formulated his famous Cooperative Principle Grice (1975) (see also Section 1 above). The impact of this work reached well beyond theoretical linguistics, also including the research traditions of computational modelling and psycholinguistics, where researchers often started from the idea that referring expressions should obey the Gricean Maxims. A key question in Grice (1975) went further, asking what ("conversational") implicatures arise in situations where speakers deviate from the Maxims.

[1] Words like "large", which are often vague in real life, do occur in the TUNA corpora, but only in situations in which there were only two, clearly distinct, sizes of objects. Thus, it was always clear whether an object was large or small.

2.2. Reference in psycholinguistics and experimental pragmatics

A wealth of work in psychologinguistics has focused on reference, often starting from the idea that reference taps into the "common ground" of speaker and hearer (Clark, 1996; van Deemter, 2016). This work has asked a range of questions about referring expressions, often centering around the degree to which speakers can be said to "design" their referring expressions so as to facilitate the process whereby the hearer can search for the referent (Clark and Murphy, 1982; Paraboni and van Deemter, 2014). A fascinating strand of this work has asked to what extent speakers and hearers are sensitive to cognitive differences between them, for example when the hearer is less well-informed than the speaker (Horton and Keysar (1996a)). The debate that has ensued around these issues is known as the *egocentricity* debate. Cognitive differences frequently give rise to referring expressions that do not contain enough information to allow the hearer to identify the referent) or, conversely, that contains more information than is necessary for this purpose. Our investigation is relevant for this debate because it will allow researchers to analyse these phenomena using a more fine-grained conceptual apparatus.

Even apart from egocentricity – that is, even pertaining to situations in which cognitive differences between speakers and hearers do not play a role – a substantial body of work has noted that over-specification is very common. Pechmann (1989); Eikmeyer and Ahlsén (1996) and Schriefers and Pechmann (1988), for example, found that attributes such as TYPE and COLOUR are ubiquitous, frequently leading to over-specification. An increasing range of studies has investigated the extent to which speakers over-specify when they refer, though typically without defining over-specification precisely and without distinguishing between different kinds of over-specification (as we will do in Section 3).

An intriguing question is whether over-specification might be useful for hearers, and *why* speakers over-specify (e.g., what is the role of audience design). Some authors have argued that over-specification can be *detrimental* to the hearer's ability to locate the target, for example *via* eye-tracking and brain scanning experiments Engelhardt et al. (2006, 2011). Paraboni et al. (2017) conducted eye-tracking studies on both atomic and relational attributes and argue that "*easily recognisable properties may facilitate identification, whereas properties that are more difficult to recognise may have the opposite effect*".

A growing body of work has shown that over-specification can be beneficial for hearers, for example when an over-specifying property taps into "prototypes" in the mind of the hearer (Levelt, 1993, Chapter 4); even when such properties are logically unnecessary for identifying the referent, they may still help the hearer. Arts et al. (2011) looked at the time readers took to identify the referent within a visual field, comparing minimal descriptions (e.g., *the button*) with over-specified descriptions that add (logically superfluous) information about the location of the referent (e.g., *the round button at the top left*. They found that some types of over-specified descriptions led to decreased identification times. Paraboni and van Deemter (2014) conducted experiments in a variety of domains and came to similar conclusions as Arts and colleagues, following up with a computational model. Recent work has confirmed these ideas and increased our understanding of the circumstances under which over-specification occurs, and the circumstances under which it is beneficial for hearers (Tourtouri et al., 2019; Rehrig et al., 2021; Rubio-Fernandez, 2021).

2.3. Reference in computational modelling

A significant strand of work in computational linguistics has focused on computationally modelling the use of referring expressions. The realisation that referring expressions sometimes need to over-specify has played a large role in this work. The fact that referring expressions also frequently *under*-specify has been systematically overlooked in this tradition. We will return to under-specification a few times, and most emphatically in Section 6.3.

Computational modelling started with algorithms that search for referring expressions that use a minimum number of attributes (Dale, 1989), that is, avoiding both under-specification and over-specification. The resulting referring expressions were found to be quite unlike the ones produced by human speakers; moreover, it was also observed that a "minimization" strategy would be so computationally time-consuming as to be psychologically implausible. Speakers can at best *approximate* the ideal of generating optimally efficient (i.e., minimal) referring expressions. To model this insight, a number of approaches were investigated, which typically select properties one by one, in a number of steps, until enough properties have been accumulated to single out the intended referent. These include "greedy" search algorithms (Dale, 1992) (which chose, at every step, the property that rules out the largest number of distractors) and the algorithm proposed in (Dale and Reiter, 1995), which selects properties following a preference order in which more highly "preferred" properties have a higher chance of getting selected for inclusion into the referring expression than less "preferred" ones. This algorithm became known as the Incremental Algorithm.

To do justice to the variations in language production that occur in any experiment, recent models have often been stochastic. In order to model the data as well as possible, these models have often assigned a prominent role to over-specification. A recent example, which builds on a long tradition of experimentation, van Gompel et al. (2019) selects attributes incrementally, in a number of steps (one step for each attribute entered into the description); at each step, the model samples an attribute from a given probability distribution, in such a way that salient attributes receive a higher probability mass. This model produces over-specified descriptions in many situations.

Another stochastic model produces referring expressions under a Bayesian framework (Frank and Goodman, 2012; Monroe and Potts, 2015). Initial versions of this "Rational Speech Act" model did not cater for over-specification (Frank and Goodman, 2012), but a recent proposal combines the Bayesian framework with "graded" semantics that seeks to model the clarity of each of the properties expressed in a given description; this model produces over-specified descriptions in certain

types of situations to help the reader identify the target referent (Degen et al., 2020). Following the same idea of helping the reader identify the target, Jara-Ettinger and Rubio-Fernandez (2021) chose to generate over-specifications differently by modelling readers' visual search of target referent directly and assumed that the aim of speakers is to minimise the time it takes to find the target.

An important aspect of the computational line of work is its emphasis on data and evaluation. To evaluate how human-like the descriptions produced by the above algorithms are, a number of data-text corpora were collected through elicitation experiments with human speakers. Examples include GRE3D3 (Dale and Viethen, 2009), TUNA (Gatt et al., 2007; Deemter et al., 2012), COCONUT (Jordan and Walker, 2005), and MAPTASK (Gupta and Stent, 2005). In the present study, we will make use of the TUNA corpora (Section 4.1).

In the Introduction to this paper, we emphasised how computational modelling in the tradition of Dale and Reiter (1995) has focused on relatively simple situations. In recent years, some computational models have targeted more complex referential situations, including reference in large and cluttered domains (Koolen et al., 2013; Paraboni and van Deemter, 2014), reference under cognitive differences between speaker and hearer (Kutlak et al., 2016), reference in dialogue (Garoufi and Koller, 2014), referring expressions that make use of vague properties (van Deemter, 2016) and expressions referring to objects in real-world images (Yu et al., 2016, 2017). A substantial amount of work on complex referential situations has modelled referring in discourse, often focusing specifically on the choice between proper names, descriptions, and pronouns (Belz et al., 2010; Kibrik et al., 2016) and in dialogue (Jordan and Walker, 2005; Villalba et al., 2017).

To get the best possible grasp of over- and under-specification, we will continue to concentrate on relatively simple situations. We will see that, even in simple situations, it is not trivial to obtain a clear perspective on the different ways in which referring expressions can deviate from the idea of referring minimally. In our Discussion section, we will discuss the relevance of the present work for such more challenging situations.

3. Varieties of referential specification: a formal account

Let us use the insights discussed in Section 1 to propose a framework for thinking about reference. Terms such as "over-specification", "superfluous", and even "wrong specification", as defined in this section, are not intended as evaluative: in some situations, it may well be a good idea to "over-specify", to express a "superfluous" property, and even – though admittedly more rarely – to specify "wrongly".

3.1. Preliminaries

As explained in the Introduction, we will make some simplifying assumptions that will facilitate our formalisation. In particular, we will assume that the speaker and listener share the same beliefs about the domain and the objects in it. (When the speaker and listener do not share these beliefs (Horton and Keysar, 1996b; Keysar et al., 2003) this complicates matters considerably.) When discussing things that can be said about a referent, we will think of these as crisp (i.e., non-vague) *properties*, each of which is either true or false of the referent. The set $\mathcal{P}$ of available properties is also considered given. Importantly, all properties that will be mentioned in our definitions below are assumed to be elements of $\mathcal{P}$. (For example, $P_1, \ldots, P_n \in \mathcal{P}$ throughout; in Definition 2, $P'_1, \ldots, P'_m \in \mathcal{P}$ likewise).[2]

We will often suppress the role of the referent r. For example, a Distinguishing Description of r will simply be called a Distinguishing Description. If $\mathcal{D}$ singles out something that is not the intended referent, we will say that it is not a Distinguishing Description but a "Wrong Description".

In Sections 3.3.2 and 3.3.4, it will be useful to group together properties into clusters, called attributes as is often done. For instance, the attribute COLOUR may have values such as red, blue, and so on. An attribute–value pair, such as ⟨COLOUR, ⟩ or ⟨TYPE, ⟩, is equivalent to a property. Because a description can sometimes express the same property twice (perhaps with different wording), a description $\mathcal{D}$ is represented by a bag[3] (i.e., multi-set) of n properties: $\mathcal{D} = \{P_1, \ldots, P_n\}$.

Last but not least, we will assume it to be a primary goal of reference to enable a hearer to identify the intended referent r in a given setting C, which consists of r itself plus a non-empty set of other objects (often called distractors, e.g., McDonald (1983)). Now the notion of a *distinguishing description* can be formally defined as follows, in which $[\![P_i]\!]$ is a set of elements that share a property P_i, i.e., the denotation or extension of P_i.

Definition 1. (Distinguishing Description). *The description $\mathcal{D} = \{P_1, \ldots, P_n\}$ is a distinguishing description of the intended referent r if it singles out r from all other elements of C. This is the case if and only if* $[\![P_1]\!] \cap \ldots \cap [\![P_n]\!] = \{r\}$.

In any Distinguishing Description, we will call an occurrence of a property *superfluous* if and only if the description would still be a Distinguishing Description if that occurrence were removed from the description.

In the relatively simple situations on which we will focus, a description cannot be successful unless it is a distinguishing description. As we noted in the Introduction, when some referents are much more salient than others, a non-distinguishing

[2] The requirement that properties are elements of $\mathcal{P}$ could have been spelt out in our definitions, but the definitions become more readable if it is used as a general convention.

[3] For instance, {*A*, *B*, *A*} is the same as {*B*, *A*, *A*} but different from {*A*, *B*}.

description may well be communicatively effective (e.g., when "the dog" is understood to be the contextually most salient dog). Similar situations arise when referential communication relies on pragmatic reasoning (Frank and Goodman, 2012; Goodman and Frank, 2016). Such situations do not occur in the empirical study of Section 4.

When the definitions are applied to a concrete text, a number of decisions need to be taken before one can say whether a given Noun Phrase is, for example, a distinguishing description. In particular, one has to decide what the available properties (the set $\mathcal{P}$) is, what the setting C is, and what elements of C each property is true of. Such decisions will be taken in Section 4, for example, where our definitions are employed to compare an English corpus with a Chinese one.

3.2. Minimal description

Early studies have often suggested that the best-referring expression for a given referent must always be the shortest possible one: a distinguishing description that uses as few properties as possible, also known as a *minimal description* (e.g., Dale (1989, 1992)). This idea can be seen as a strict interpretation of the Maxim of Quantity.

Definition 2. (Minimal Description). *A set of property occurrences $\mathcal{D} = \{P_1, \ldots, P_n\}$ is a minimal description if and only if it is a Distinguishing Description and there is no Distinguishing Description $\mathcal{D}' = \{P'_1, \ldots, P'_m\}$ such that $m < n$, that is, $|\mathcal{D}'| < |\mathcal{D}|$.*

Here, $|\mathcal{D}|$ is the size of $\mathcal{D}$, that is, the number of property occurrences in $\mathcal{D}$. It is easy to see that, in one and the same situation, a referent may have more than one minimal description. Note that we have defined minimality on the basis of the number of properties in the description. Whether the resulting notion of minimality is more suitable for capturing linguistic generalisations than notions of minimality defined in terms of, for example, the number of words or syllables in the surface form of a description is a question for future research.

3.3. Over-specification

Previous studies (Engelhardt et al., 2006, 2011; Koolen et al., 2011) motivate their understanding of over-specification on the basis of the second part of the Gricean Maxim of Quantity: a referring expression is over-specified if it is more informative than is necessary for successful communication. This clearly covers situations in which a referring expression includes non-required properties while managing to identify the referent. However, as discussed in Section 1, there are some interesting distinctions that this definition does not make because, as illustrated in example (1), a description without superfluous properties may nonetheless not be minimal.

Definition 3. (Over-specified Description). *A set of property occurrences $\mathcal{D} = \{P_1, \ldots, P_n\}$ is an Over-specified Description if and only if it is a Distinguishing Description and it is not a Minimal Description.*

We will now sub-categorise the class of Over-specified Descriptions.

3.3.1. Numerical over-specification

Fig. 2 shows a scene from the MTUNA corpus. When subjects referred to the chair in the window, they could say any of the referring expressions in (2), where the "MD" mark indicates the current description is a minimal description:

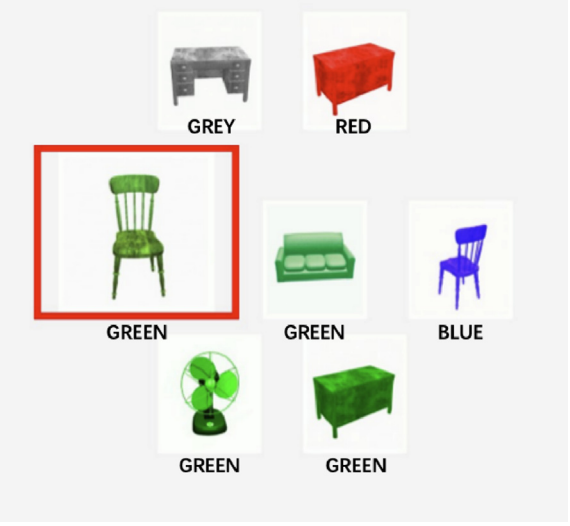
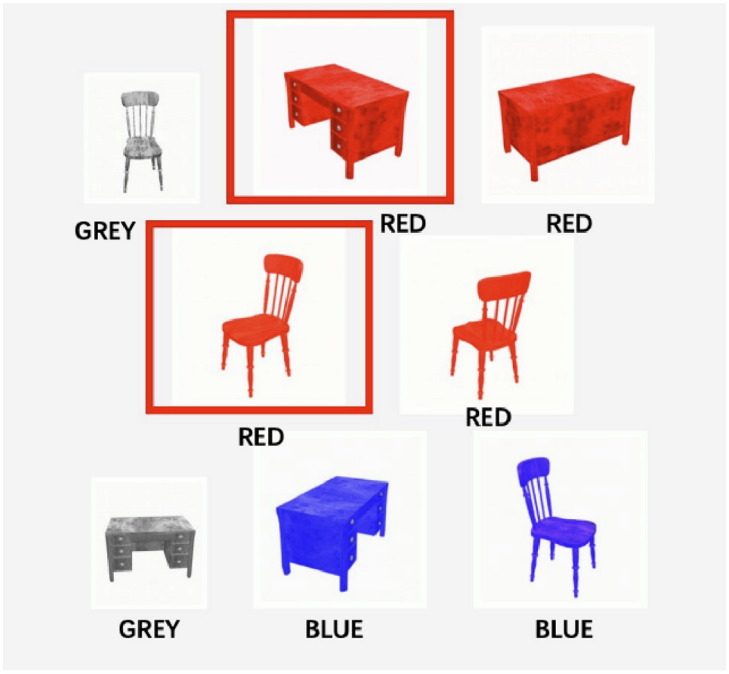
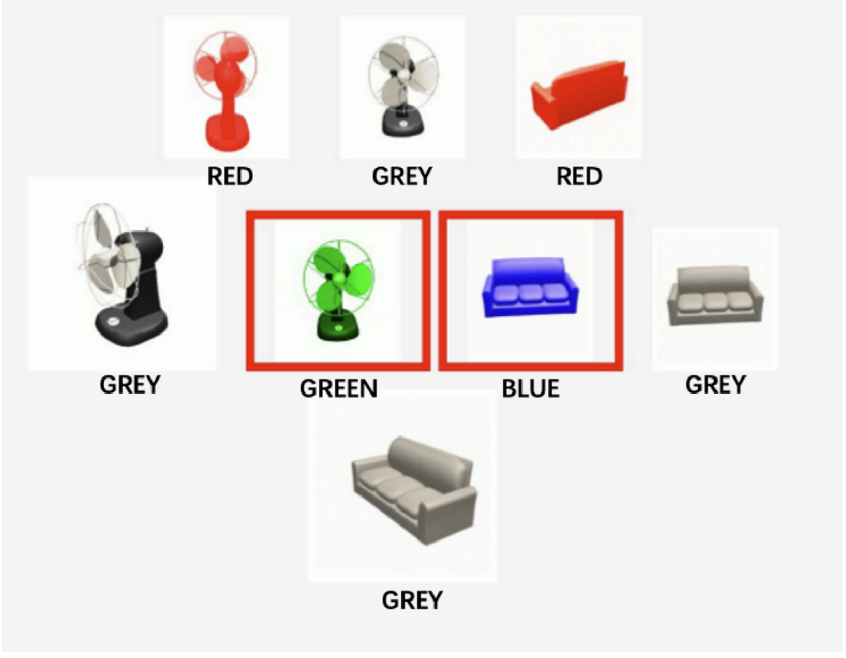


Fig. 2. A scene from the MTUNA corpus, where the object in the window is the target object.



(a)



(b)

Fig. 3. Two scenes from the MTUNA corpus, each of which is a scene asking subjects to produce referring expressions referring to a set of two target referents.

(2) a. the large one (MD)
b. the large green one
c. the green chair

As there is only one large object in the scene, using only one property: ⟨SIZE, large⟩, is sufficient for successful communication, as in (2-a). However, except over-specifications, such as (2-b) whose property ⟨COLOUR, green⟩ removable

without breaking the communication successful, there are also descriptions like (2-c), which uses two properties: ⟨COLOUR, green⟩ and ⟨TYPE, chair⟩. Clearly, description (2-c) uses more properties than the minimal description. But, simultaneously, none of its properties is superfluous.

We call descriptions like (2-c) Numerical over-specifications. In descriptions of this kind, no property is superfluous (unlike (2-b), where "green" could be removed) yet it is possible to construct a *shorter* referring expression by replacing a set of properties in the expression by a smaller set of properties where the result is still a distinguishing description:

Definition 4. (Numerical Over-specification). *The description* $\mathcal{D} = \{P_1, \ldots, P_n\}$ *is a numerical over-specification if and only if* $\mathcal{D}$ *is a Distinguishing Description and there is no* $P \in \mathcal{D}$ *such that* $\bigcap_{P_j \in \mathcal{D} - \{P\}} [\![P_j]\!] = \{r\}$*, but the number of attributes n is greater than that of a minimal description of r.*

3.3.2. Nominal over-specification

The special status of the TYPE attribute comes from a long tradition of psycholinguistic work, summarised well in Levelt (1993, Chapter 4), based on the idea that English speakers tend to include a head noun in their referring expressions. This idea was combined with the idea that head nouns classify things into broad classes ("types") of objects (e.g. (Dale and Reiter, 1995),), thus differentiating the TYPE attribute from other attributes. TYPEs are also known as *categories* and a large body of theory has arisen about the role of types in language and thought (Rosch et al., 1976). For the scene in Fig. 2, we call descriptions like:

(3) the large chair

where there is a superfluous TYPE attribute while none of its other attributes is superfluous attributes, as *nominal over-specifications*. Formally:

Definition 5. (Nominal Over-specification). *A Nominal Over-specification is a set of property occurrences* $\{P_1, .., P_n\}$ *in which at least one of* $P_1, .., P_n$*, say* P_i*, is a TYPE, and* $\{P_1, .., P_n\} - \{P_i\}$ *is a Distinguishing Description, but for every* $j \neq i$*,* $\{P_1, .., P_n\} - \{P_j\}$ *is not a Distinguishing Description.*

Note that we don't need to require explicitly that $\{P_1, .., P_n\}$ is distinguishing because this follows from the requirement that $\{P_1, .., P_n\} - \{P_i\}$ is a Distinguishing Description.

3.3.3. Duplicate-attribute over-specification

Similar to the description (1-f), regarding the target object in Fig. 2, one could say:

(4) the green chair that has the same colour as the fan

This description contains repeated use of the same attribute COLOUR (i.e., both the phrase *green* and the phrase *has the same colour as the fan* are talking about colour). We call this a *Duplicate-Attribute Over-specification*.

Definition 6. (Duplicate-Attribute Over-specification). *A description* $\mathcal{D} = \{P_1, \ldots, P_n\}$ *is a duplicate-attribute over-specification if and only if there exist two property occurrences* $P_i, P_j \in \mathcal{D}$ *such that* $P_i = P_j$*, and* $\bigcap_{P_k \in \mathcal{D} - \{P_i\}} [\![P_k]\!] = \{r\}$.

The clause "$P_i = P_j$" means that P_i and P_j express the same value of the same attribute. Note that this does not preclude considerable variations in surface form. For example, the *green* and the *has the same colour as the fan* express the same property. Other examples in the TUNA corpora include sofa vs. settee, male vs. man, small vs. little, and so on.

3.3.4. Real over-specification

Now let us turn to the kind of over-specification that was covered by the bulk of previous studies, namely (what we will call) *Real Over-specification*. For instance, for the target referent in Fig. 2, the description (2-b) is a real over-specification. Real over-specification does not overlap with nominal over-specification. In other words, if a description has superfluous properties in addition to a superfluous TYPE, it is a real over-specification no more. Concretely, a more formal definition can be written as:

Definition 7. (Real Over-specification). *A description* $\mathcal{D} = \{P_1, \ldots, P_n\}$ *is defined as a real over-specification if at least one of the* $P \in \mathcal{D}$ *is* $P \neq$ TYPE *and such that* $\bigcap_{P_j \in \mathcal{D} - \{P\}} [\![P_j]\!] = \{r\}$.

It could be argued that there exists another special type of over-specification, where the value of an attribute is more specific than necessary. In TUNA, such over-specification occurs frequently in the TYPE attribute in the people sub-corpus (see Section 4.1 for an introduction of the TUNA corpus), for example when the word *scientist* is used even though the word *person* would have been sufficient. This situation could be modelled by saying that *scientist* is a sub-type of *person*. This phenomenon might be called *Choice-of-Value Over-specification*.

We have chosen not to follow this approach because it would create a systematic ambiguity. An over-specified description could be turned into a minimal description in different ways: by removing a property (e.g. removing a property like "wears glasses"), or by replacing a property by a more general one (e.g. replacing scientist by person); the former would make it a real over-specification, but the latter would make it a choice-of-values over-specification.

To acknowledge the fact that, in the situation above, "scientist" is over-specified, we proceed as follows. A sub-type is interpreted as introducing new attributes to its parent type, i.e., dividing a single attribute into multiple attributes. For example, the word *scientist* expresses both TYPE and JOB.

3.4. Under-specification

Under-specification is the flip-side of over-specification. In the simple situations on which we focus, under-specification is about descriptions that do not successfully single out the target referent. It breaks the first principle of Gricean Quantity, because the speaker did not make the contribution as informative as required. For Fig. 2, the description (5-a) cannot help the reader to successfully identify the intended referent as there are two chairs in the scene.

(5) a. the chair
b. the large one (MD)
c. the green chair.
d. the large chair

This kind of specifications be defined as follow:

Definition 8. (Under-specification). *If, for a description* $\mathcal{D} = \{P_1, \ldots, P_n\}$, *there exists a real super-set A of r ((i.e.,* $\{r\} \subsetneq A$*)) such that* $\bigcap_{P_j \in \mathcal{D}}[\![P_j]\!] = A$, *then we call* $\mathcal{D}$ *an Under-specification.*

Analogous to real over-specification, if a description contains no superfluous property, but one of its properties is an attribute whose value is not specific enough, then this could also be seen as a special type of under-specification, namely, *Choice-of-Value Under-specification*. In example (6), if there are two chairs in a scene, where one is blue and the other is black, then compared to the minimal description (6-b), (6-a) is not specific enough to single out the referent.

(6) a. the dark-coloured chair
b. the blue chair (MD)

A similar phenomenon can occur when referring to multiple objects (Section 5.4). In MTUNA, for the scene 3(a), we found the following referring expressions (translated from Mandarin):

(7) a. the objects that are seen from the side
b. the red objects that are seen from the side.
c. the left facing objects (MD)

If *seen from the side* means "facing left or right", then (7-a) is choice-of-value under-specified, compared to the minimal description (7-c). To avoid such under-specification there are two alternatives: 1) making the property more specific (e.g., *left facing*) and constructing a minimal description; 2) adding a COLOUR property with the value of *red*, resulting in a numerical over-specification (7-b). It is thus unclear how the description should be analysed: as a regular under-specification, or as a choice-of-value under-specification.

The same problem occurs in scene 3(b), where one could say:

(8) a. the green one and the blue one (MD)
b. the front-facing coloured objects (MD).
c. the coloured objects

Similar to (7-a), for (8-c), one could either replace the word *coloured* with specific colour terms for the two objects respectively and distribute them into two clauses, or one could add the ORIENTATION of these two objects as in (8-b). In what follows, we will not make use of the notion of Choice-of-Value Under-specification.

3.4.1. Mixed description

The literature has tended to focus on situations in which a description is either over-specified or under-specified, or minimally specified. Logically, however, there are other possibilities, and these are also encountered in real life.

One example of such a case is what we call a *Mixed Description*. A Mixed Description is an Under-specified Description from which one or more property occurrences can be removed without changing the extension of the description. One might say that such a description is both over-specified and under-specified. More precisely:

Definition 9. (Mixed Description). *A description* $\mathcal{D}$ *is a Mixed Description if and only if it is an Under-specified Description and there exists a property occurrence* P_i *in the description such that* $\bigcap_{P_k \in \mathcal{D}-\{P_i\}}[\![P_k]\!] = \bigcap_{P_k \in \mathcal{D}}[\![P_k]\!]$.

For example, given the scene depicted in Fig. 4, the under-specification (9-a) describes both the SIZE and the COLOUR of the referent. However, all small objects in the scene are green, which suggests that the use of COLOUR does not add any information. In other words, COLOUR is superfluous in this under-specification. By contrast, both ORIENTATION and TYPE in (9-b) help to single out the referent, hence it's not a mixed description but a pure under-specification, a notion we will presently define.

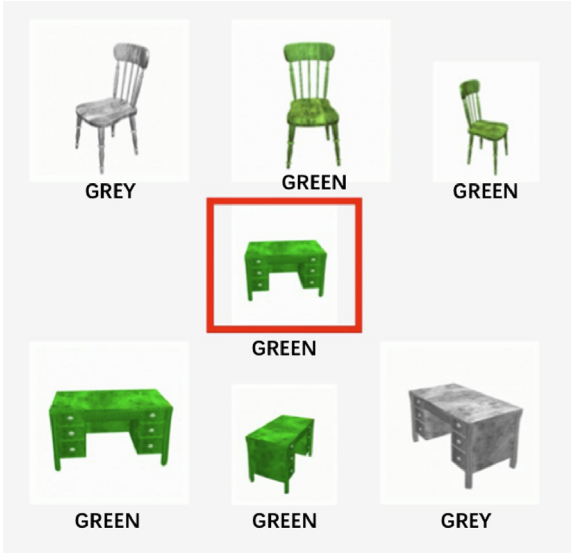


Fig. 4. A scene from the MTUNA corpus.

(9) a. the green small desk
b. the front-facing desk

3.4.2. Pure under-specification

Pure under-specifications are under-specifications that are not mixed:

Definition 10. (Pure Under-specification). *A set of property occurrences* $\mathcal{D} = \{P_1, \ldots, P_n\}$ *is a Pure Under-specified Description if and only if it is an Under-specification and it is not a Mixed Description.*

3.5. Wrong description

In some cases, a description is simply *wrong*, i.e., describing the target referent incorrectly. In the TUNA corpora, approximately 3–4% of all descriptions fall into this category; in everyday conversation, the frequency of wrong descriptions might be even higher.

Definition 11. (Wrong Specification). *A Wrong Specification is a set of property occurrences* $\{P_1, .., P_n\}$ *in which at least one of* $P_1, .., P_n$, *say* P_i, *is not true of the intended referent r, that is,* $r \notin [\![P_i]\!]$.

It follows that when $\{P_1, .., P_n\}$ is a Wrong Specification, then $r \notin [\![P_1]\!] \cap \ldots \cap [\![P_n]\!]$. Note that, just like underspecified descriptions, wrong descriptions may not always be unsuccessful, for example because hearers may be able to dismiss the incorrect information (e.g. when the description is overspecified) or "correct" the information in light of the situation at hand.[4]

3.6. Description basis

So far, we have introduced various kinds of over-specification. By virtue of their definitions, minimal descriptions and numerical over-specifications do not contain any superfluous properties. Therefore, they can serve as a "basis" for various (other) types of over-specification. Formally, given an over-specified description X that has r as its intended referent, where this description expresses property occurrences $\{P_1, .., P_n\}$, then this description is considered to be "built around" a minimal description (or a numerical over-specification) if there exists a proper subset X_b of X of $\{P_1, .., P_n\}$ such that X is a minimal description (or a numerical over-specification) of r. X_b will be called a description basis of X.

The idea of a minimal description "basis" around which a description can be seen as having been built up will be useful in Section 4.

[4] A famous theoretical study of wrong descriptions is Donnellan's study of sentences like "The man with the Martini is the murderer of Smith", when the criminal in question in fact drinks something other than Martini citetdonnellan.

3.7. The logic of reference

Given the definitions above, a number of things follow. Here we state some of the more important relations between classes of descriptions and we visualise these in a graph. Since most of these consequences of our definitions are fairly immediate, most theorems will be stated without formal proof.

Theorem 1. *All minimal descriptions, real over-specifications, numerical over-specifications, nominal over-specifications, and duplicate-attribute over-specifications are distinguishing descriptions.*

Theorem 2. *A distinguishing description cannot be a mixed description, a pure under-specification, or a wrong description.*

Theorem 3. *Each of the following classes is mutually exclusive: minimal descriptions, real over-specifications, numerical over-specifications, nominal over-specifications, mixed descriptions, pure under-specifications, and wrong descriptions.*

Theorem 4. *Each duplicate-attribution description is either a real over-specification or a nominal over-specification.*

Fig. 5 describes the relationship between each type of specification, in which each block represents a category we have introduced in this paper. Consider theorem 4 for instance. Suppose we have a duplicate-attribute over-specification $\mathcal{D} = \{P_1, \ldots, P_n\}$ in which there are $m(0 < m < n)$ duplicated properties (see Definition 6 for the definition of duplicated properties), represented as $\mathcal{D}_{dup}$ ($\mathcal{D}_{dup} \subseteq \mathcal{D}$). If all properties in $\mathcal{D}_{dup}$ are TYPEs, then $\mathcal{D}$ is a nominal over-specification (because TYPE is the only superfluous attribute in $\mathcal{D}$). Otherwise, it is a real over-specification (because $\mathcal{D}$ contains a superfluous non-TYPE property). The other theorems can be proven using similar reasoning.

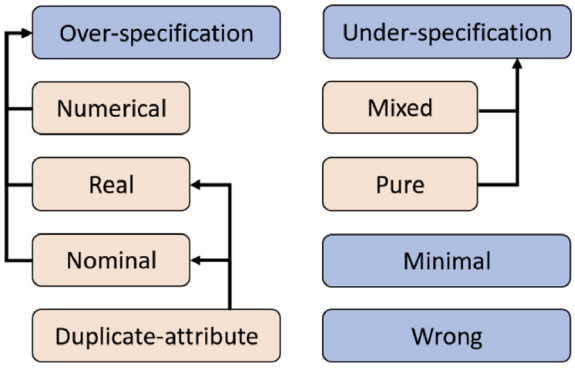


Fig. 5. Diagram of relationships between each type of specification. $A \rightarrow B$ means A is a kind of B.

4. The framework applied

In order to demonstrate what happens when our new framework is employed to analyse referring expressions produced by human speakers, we examine some corpora that were designed specifically to gain insight in reference: the TUNA corpora (sometimes referred to collectively as *TUNA).

We will first introduce the MTUNA and ETUNA experiments. Second, we describe how we annotated these corpora. Lastly, we describe how these newly annotated corpora were analysed, and what were some of the findings that this gave rise to.

4.1. Introducing and annotating the TUNA corpora

*TUNA is a series of experiments in which human speakers were asked to refer to a sequence of objects under carefully controlled conditions. This was done with the immediate goal of comparing the output of a computational model (in which referring expressions are generated automatically) against a body of referring expressions produced by human speakers. This was done for a number of languages, including English (Gatt et al., 2007; Deemter et al., 2012), Mandarin Chinese (van Deemter et al., 2017), Dutch (Koolen and Krahmer, 2010) and German (Howcroft et al., 2017). The procedure was always to ask participants to produce expressions that refer to whatever object appeared in a red window on a computer screen, among a set of "distractor" objects (Fig. 2) outside that window. Details vary between the different TUNA experiments. This setup, honed in the context of computational modelling, lends itself well for illustrating the definitions of Section 3 and the annotation scheme of Appendix A.

In this paper, we focus on the Mandarin Chinese version of TUNA, that is, on MTUNA, so we will say a bit more about the way in which MTUNA was conducted. The stimuli used by MTUNA are inherited from the Dutch TUNA (Koolen and Krahmer, 2010; Koolen et al., 2011), which contained 40 trials. Four additional trials were used as a part of the instructions to subjects (this includes Fig. 2). The experiment used scenes from 2 domains: the furniture domain and the people domain. The furniture

domain uses artificial pictures as in Figs. 2 and 3. The people domain uses black-and-white photographs of male mathematicians as in Fig. 6. Each domain has 20 trials, half of which have a singular target (i.e., one piece of furniture or one person) and the other half have two targets. Unlike other TUNAS, where subjects were always asked the same question (namely *Which object/objects appears/appear in a red window?*), the organisers of the MTUNA experiment were curious to see whether it made a difference whether the referring expression occurs in subject or object position. Therefore, subjects were asked to write referring expressions on a series of blanks in either of the following patterns:

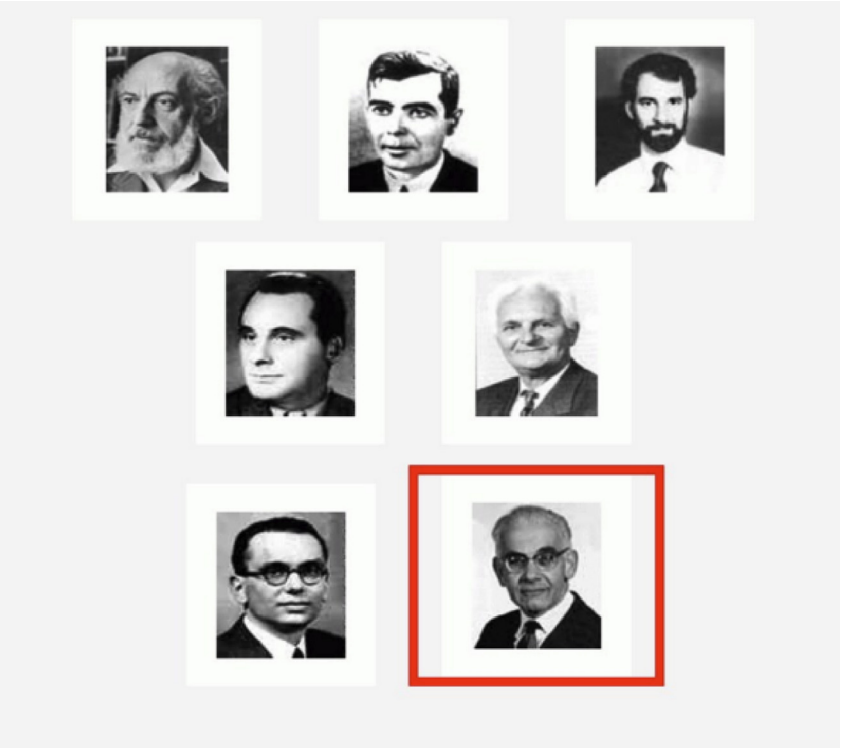


Fig. 6. A scene of from people domain of the MTUNA corpus.

(10) a. ____zai hongse fangkuang zhong
'Please complete the sentence: ____is in the red frame(s)'
b. Hongse fangkuang zhong de shi ____
'What's in the red frame is ____'

where (a) asked subjects to place the referring expression in the subject position while (b) asked to place it in the object position.

We annotated the MTUNA corpus following the practice described in Appendix A, but we limited the corpus in a number of ways. Firstly, we focused only on the part of the corpus in which the referent was one object (rather than two). We also excluded situations in which minimal descriptions or numerical over-specifications use TYPE. In other words, we focused on situations where TYPE is always superfluous. The resulting sub-corpus contains 7 trials in the furniture domain and 10 trials in the people domain. Secondly, we also annotated the English ETUNA corpus in order to compare it with Chinese MTUNA. To allow a fair comparison between the two corpora, we focused our analysis on ETUNA trials where participants did not use *location* (e.g. "the desk *in the top left*"). Finally, we only use trials in which the two corpora contain exactly the same scene, which yielded 7 trials in the furniture domain and 6 trials in the people domain. After annotation, we counted the frequencies of each type of specification and recorded them in Table 1.

Table 1
The number of singular referring expressions in each category. **total** is the number of valid descriptions. **mini.** is minimal description, **real** is real over-specification, **nom.** is nominal over-specification, **num.** is numerical over-specification, **dup.** is duplicate-attribute over-specification, **mix.** is mixed description, **pure** is pure under-specification, **over** is over-specification, and **under** is under-specification. Note that when counting the total number of over-specifications, we need to exclude the number of duplicate-attribute over-specification.

			over				under	
	total	mini.	real	nom.	num.	(dup.)	mix.	pure
Furniture	252	9	84	104	0	3	11	44
People	361	17	217	69	13	2	43	2

4.2. Analysis of MTUNA

The people domain is more complex than the furniture domain along two dimensions: 1) scenes in the people domain use real photographs of people, which allows a wider variety of attributes to choose from compared to the artificial pictures in the furniture domain. 2) Since all the referents in the people domain are male scientists, these are arguably more perceptually similar to each other than the referents in the furniture domain. Accordingly, participants over-specify more frequently in the people domain than in the furniture domain (cf. van der Sluis and Krahmer (2007)).

In a closely related study in this journal, Koolen et al. (2011) hypothesised that speakers tend to convey more information in the people domain than in the furniture domain. To test this idea, they compared the number of superfluous attributes and found that, on average, referring expressions in the people domain contain more superfluous attributes. Note that a higher number of superfluous attributes does not necessarily result in the more frequent use of over-specified descriptions, for example, because of the existence of numerical over-specifications (which over over-specified without containing any superfluous properties), and because of the existence of mixed specifications (which are under-specifications that contain some redundancy at the same time). Furthermore, Koolen and colleagues didn't differentiate TYPE from other attributes when counting the number of superfluous attributes.

To therefore test the idea that there is a higher proportion of over-specifications in the people domain than in the furniture domain, we counted the number of over-specifications in MTUNA. We counted as over-specifications all descriptions that are annotated as real, nominal, or numerical over-specification. This resulted in 188 and 299 over-specifications in the furniture sub-corpus and the people sub-corpus, respectively (see Table 1).

How should *non*-over-specification be counted? With our annotation scheme in hand (see Appendix A), different approaches are possible. One option is to count all valid descriptions that are not annotated as over-specification, which yields 64 and 62 descriptions, respectively. Using a Chi-square test with Yates correction, we were able to confirm the above hypothesis with moderate significance, $\chi^2(1, N = 613) = 6.1441$, $p = .0132$.

However, this way of counting may overlook over-specified properties in mixed descriptions. Therefore, we also tested the above hypothesis (i.e., that there are more over-specifications in the people domain than in the furniture domain), omitting all descriptions that do not result in successful communication (i.e., under-specification) and, this time, we found no significant difference, $\chi^2(1, N = 513) = 0.166$, $p = .6837$. This result is not surprising once one realises that *nearly all* successful descriptions in both the furniture and the people corpora were over-specifications: recall that speakers tend to include a TYPE no matter whether it contributes to distinguishing the target and this biases the analysis (e.g., Levelt (1993)). Therefore, to obtain a more insightful analysis of how domain difficulty influences over-specification, we focused on those over-specifications that are not nominal over-specification, hypothesising that there are more of these over-specifications in the people domain than in the furniture domain. We, therefore, summed up all the real, numerical and duplicate-attribute over-specifications, resulting in the numbers 84 and 230, for the furniture and the people sub-corpus, respectively. We once again tested the hypothesis and found that this time, it was confirmed with high significance, $\chi^2(1, N = 513) = 46.4435$, $p < .0001$.

While a post-hoc analysis of this kind − where different definitions of key phenomena are attempted − has to be treated with some caution, these results at least demonstrate how different insights can be gleaned depending on *what kind* of over-specification one wants to focus on.

Another type of analysis concerns other end of the specification spectrum, namely under-specification. We investigated whether domain difficulty influences the use of under-specification, i.e., *are speakers less likely to single out the target object when the domain is more complex?* To find out, we counted the descriptions that were annotated as under-specification in MTUNA and obtained 55 and 45 under-specifications for furniture and people corpus, respectively. Surprisingly, our hypothesis is not true; in fact, the situation is the other way around, $\chi^2(1, N = 613) = 9.5237$, $p = .0020$. Understanding why this happened is a topic for further research.

Note, furthermore, that the results (15.74% of descriptions in MTUNA are under-specifications) show that the role of under-specifications cannot be ignored when analysing referring expressions. We will return to this important issue at the end of this section and in the Discussion section.

Our annotation scheme allows us to look in detail at specific attributes, such as TYPE. As we have seen in §4.1, in the people domain, TYPE has only one possible value (i.e., man) whereas, in the furniture domain, there are multiple alternatives (e.g., table, chair or sofa). In other words, the "attribute complexity" of TYPE in the furniture domain is higher than that in the people domain (Lv, 1979). suggested that the head of a Chinese noun phrase can be omitted if it is the only possible head noun given the context. It, therefore, seemed reasonable to us to expect that speakers are more likely to add a redundant TYPE in the furniture domain than in the people domain. Table 2 shows that 91.29% of referring expressions in the furniture domain and 77.90% in the people domain use TYPE. This confirmed our expectation ($\chi^2(1, N = 614) = 18.2171$, $p < .0001$). Previous

Table 2
The number of successful communications with or without superfluous TYPE in furniture and people domain respectively.

	w/superfluous TYPE	w/o superfluous TYPE
Furniture	230	22
People	282	80

research has indicated that TYPE is more likely to be dropped when referring to animate than to inanimate referents (Fukumura and van Gompel, 2011), which might offer an additional explanation for our finding.

4.3. Comparing MTUNA with ETUNA

To show how the framework contributes to the comparison between two languages that use reference very differently, we conducted a comparison between referring expressions in Mandarin Chinese and English in successful communications (Table 3), i.e., disregarding under-specification.[5] Let's compare these results with our earlier findings, which concerned the two corpora as a whole. For the Chinese, our earlier conclusions still hold: (1) there are more over-specifications in the people domain than in the furniture domain, $\chi^2(1, N = 470) = 6.80$, $p < .01$; (2) in successful communication, there are more real over-specifications in the people domain than in the furniture domain, $\chi^2(1, N = 396) = 39.72$, $p < .0001$; (3) There are more under-specifications in the furniture domain than in the people domain, $\chi^2(1, N = 470) = 15.14$, $p = .0001$; (4) There are more TYPEless descriptions in the people domain than in the furniture domain, $\chi^2(1, N = 483) = 24.83$, $p < .0001$.

Table 3
The number of singular referring expressions that fall in each category in MTUNA and ETUNA respectively.

				over				under	
	domain	total	mini.	real	nom.	num.	(dup.)	mix.	pure
MTUNA	**furniture**	252	9	84	104	0	3	11	44
	people	218	15	145	37	2	2	18	1
ETUNA	**furniture**	156	1	59	62	0	0	6	28
	people	132	3	75	47	0	1	7	0

Moving to the ETUNA corpus, some of the above conclusions regarding Mandarin Chinese carry over. In particular, (1) There are more over-specifications in the people domain than in the furniture domain, $\chi^2(1, N = 288) = 11.97$, $p < .001$; (2) There are more under-specifications in the furniture domain than in the people domain, $\chi^2(1, N = 290) = 13.63$, $p = .0002$.

However, some of our conclusions do not carry over. For instance, in ETUNA we did not find statistically reliable confirmation for the idea that there should be more real over-specifications in the people domain than in the furniture domain ($\chi^2(1, N = 247) = 3.37$, $p = .0664$). One possible explanation is that English speakers tend to over-specify regardless of how complex the domain is.

We also did not find any significant difference between the use of English TYPEless descriptions in the furniture domain and in the people domain (Table 4, $\chi^2(1, N = 244) = 0.3496$, $p = .5544$). We further tested the difference of the impact of language on the use of TYPEless descriptions. We found that Mandarin speakers use significant more TYPEless descriptions than English speakers ($\chi^2(1, N = 781) = 34.7993$, $p < .0001$).

Table 4
The number of successful communications with or without superfluous TYPE in each domain in MTUNA and ETUNA respectively.

	domain	w/superfluous TYPE	w/o superfluous TYPE
MTUNA	**furniture**	241	23
	people	163	56
ETUNA	**furniture**	158	3
	people	132	5

It has been suggested that East Asian languages handle the trade-off between brevity and clarity differently to those of Western Europe, (e.g., Newnham (1971), and Huang (1984)), with the former allegedly leaning more towards brevity, and relying more on communicative context for disambiguation. If the theory is correct, one might expect that *there are more over-specifications in* ETUNA *than in* MTUNA, and that *there are more under-specifications in* MTUNA *than in* ETUNA. However, we did not find any reliable difference in the use of over-specification, $\chi^2(1, N = 758) = 3.19$, $p = .743$, the use of real over-specifications out of successful communications, $\chi^2(1, N = 643) = 1.03$, $p = .3095$, and on the use of under-specification, $\chi^2(1, N = 758) = 0.31$, $p = .5742$.

4.4. Further observations

Several further observations are worth reporting.

4.4.1. Duplicate-attribute over-specification

In the part of the MTUNA corpus that contains references to single objects, rather than pairs of objects, we observed only 5 duplicate-attribute over-specifications. In all these cases, duplication happened when a speaker used one single word to express multiple attributes, as in the following MTUNA description:

[5] Recall that, in the corpora at hand, under-specification implies unsuccessful communication.

(11) nianlao de zhangzhe
'the old old person'

where the word "zhangzhe" expresses both the AGE (old) and the TYPE (person) of the target. Given the small number of cases, this observation needs to be handled with caution, of course.

4.4.2. Numerical over-specification

We were curious about the number of specifications that use a numerical over-specification as their description basis. Out of 17 trials of MTUNA, 7 allow numerical over-specification, all of which are in the people domain (which explains the fact that no numerical over-specification is found in the furniture domain, Table 1). 83 out of 260 referring expressions (approximately 31.92%) are either numerical over-specifications or over-specifications that are built around numerical over-specified.

4.4.3. Under-specification

As we can see from Table 1 and Table 3, 15.74% and 14.23% descriptions in MTUNA and ETUNA are under-specifications, which is much higher than the previously reported proportions (e.g., Koolen et al. (2011) reported 5%). The TUNA domains are sufficiently simple — without any evident differences in salience between objects, for example — that it is fair to assume that, in these domains, a description was successful if and only of it was distinguishing. Consequently, approximately 15% of descriptions did not result in successful communication. In light of earlier discussions of under-specification (e.g. Koolen et al. (2011)), this finding was surprising, which is why we discuss it further in Section 5.2.

4.4.4. Grammatical Role

As we have seen in Section 4.1, MTUNA differentiated between the subject and object position of a referring expression (van Deemter et al., 2017). We expected that this difference in syntactic position would not affect over-specification. But, intriguingly, we found that descriptions in subject position contain many more under-specifications and far fewer real over-specifications. We consider the causes of this finding to be a matter for further research. Definiteness may play a role here given that, in Mandarin, the subject position (and other pre-verbal positions) favour definiteness (Chao, 1965).

5. Discussion

We summarise the benefits of our outlook on specification from a theoretical perspective and from a perspective of analysing a corpus of referring expressions (Section 5.1). We then discuss implications for computational models of referring (Section 5.2). We also discuss implications for our understanding of reference in complex situations (Section 5.3) and how our study can be extended to handle referring expressions that refer to sets (Section 5.4).

5.1. Lessons for linguists

We have argued that the literature on the pragmatics of referring expressions has often been imprecise because key terms are left undefined (Section 1). It has also been rather narrow, because the bulk of attention has gone to over-specification alone, and because important distinctions have not been made. To remedy these shortcomings, our new perspective proposed precise definitions of a number of key concepts and points out that each of them is not one phenomenon but many. Among other distinctions, we found that over-specification — one of the key concepts in this area — can cover different phenomena, which we dubbed nominal, numerical, and real over-specification (Section 3.3).

The benefits of our perspective for the working linguist were illustrated by comparing the referring expressions in Chinese and English by studying a bi-lingual data-text corpus. On the one hand, we found that both languages contain similar amounts of both over-specification and under-specification (see Section 5.3); on the other hand, we found that one specific type of description, namely nominal over-specifications, was substantially more frequent in Chinese than in English. Such a result would not have been possible without making fine distinctions within the conventional category of over-specification.

We believe that our perspective can have important implications for the direction of research in theoretical and experimental pragmatics. For example, for each of the sub-classes of over-specification (real over-specification, nominal over-specification, numerical over-specification, etc.): (1) The likelihood of occurrence may be different. For example, as we have seen, previous research has shown that nominal over-specification is particularly frequent. (2) The effect on hearers may be different; for example, the more frequent types of over-specification could slow down readers less than the less frequent ones, for example, because they are less surprising; and (3) The question of what implicatures are triggered may have different answers; for example, in relation to Fig. 1, in the Introduction, we conjecture that utterance 1(b), "The green chair", does not trigger any false implicatures despite using more information than is necessary to identify an object that could have been singled out equally effectively by saying "the large one". It would be interesting to follow up on earlier experimental studies of the effect of over-specification on hearers Engelhardt et al. (2006, 2011); Paraboni and van Deemter (2014) by investigating whether numerical and nominal over-specifications help readers or slow readers down, compared to a real over-specification that uses the same number of properties.

Under-specification appears in a new light as well, and not only because our work suggests that this phenomenon is more frequent than previously thought (Section 6.3). As discussed in Section 3.4, under-specification should be seen as covering

different phenomena depending on whether the expression in question is referentially ambiguous in the context in which it is used. The recent literature suggests that there can be genuine under-specifications that do not result in referential ambiguity because the ambiguity can be resolved by pragmatic reasoning (Frank and Goodman, 2012) (Cf., Section 3.1). Our investigation identified a further, previously overlooked, type of under-specification, which we called mixed under-specification. Descriptions of this type meet the definition of under-specification, but they use redundant properties also. More research is needed to find out under what circumstances descriptions of this kind are used, and how they are interpreted.

5.2. Lessons for models of reference

The production of referring expressions has long been a focus of work in Natural Language Generation (see e.g. Krahmer and van Deemter (2012)) where highly explicit computational models of referring are constructed, which seek to reproduce the referential choices made by speakers. Our findings have implications for this line of work as well. For example, our finding that Mandarin speakers are less likely to use TYPE suggests that if one were to adapt referring expression generation algorithms to Mandarin Chinese, TYPE may not deserve the special role allotted to it in models like the ones discussed in the literature (e.g., Dale and Reiter (1995)).

Remarkably, we found substantial proportions of under-specifications in both the Mandarin and the English corpus, as much as 15–20%. The computational modelling literature (Krahmer and van Deemter, 2012) has been largely silent on under-specification, and when under-specified descriptions are mentioned, their frequency is reported at around 5% (Matthews et al., 2006; Pechmann, 1989; Ferreira et al., 2005; Koolen et al., 2011). Accordingly, most language production models in this area, from Dale (1989); Dale and Reiter (1995) to van Gompel et al. (2019) do not produce *any* descriptions that do not contain enough information to permit identification (by the hearer) of the intended referent (except those following the Bayesian framework, e.g., Degen et al., 2020); this is remarkable given their stated aim of reproducing human referential behaviour.

Finally, we believe that distinctions between different types of specification should be taken into account in the experimental *evaluation* of algorithms that model human reference production (e.g. Deemter et al. (2012) and Gatt and Belz (2009)), because this will give a much more fine-grained and informative picture of the extent to which a generation algorithm is able to emulate human language production. This is for the following reason.

When referring expression generation algorithms are evaluated, they are typically evaluated in terms of metrics such as DICE, (Dice, 1945), which are based on a simple comparison between two sets, in terms of the proportion of elements they have in common. In this case, the sets are D_H, containing the attributes expressed in the description produced by human subjects and D_A, the set of attributes chosen by a generation algorithm:

$$\mathrm{DICE}(D_H, D_A) = \frac{2 \times |D_H \cap D_A|}{|D_H| + |D_A|} \tag{1}$$

Because DICE only looks at the proportion of properties that end up in both D_H and D_A, without comparing their communicative effects, DICE only bears a very feeble relation to the notions of over-specification and under-specification. For instance, it is easy to see that two referring expressions can have the same DICE score even though one is an under-specification (or a wrong specification) whereas the other is an over-specification. In the computational linguistics community, there is a growing awareness that many standard evaluation metrics are too crude to offer much insight (e.g. Reiter (2018) for a discussion of the BLEU metric in this light). A variant of DICE that boosts the score of a generated description if it belongs to the same specification class as the gold-standard description in a corpus against which it is being compared (e.g. because both descriptions are over-specified) would be a step in the right direction.

5.3. Generalising to more complex situations

In this paper so far, we have disregarded the role of linguistic context in the interpretation of a referring expression; we have also disregarded the fact that the words in a referring expression can be vague. In what follows, we discuss how our ideas may be applied to the interpretation of referring expressions in discourse context, and to the interpretation of referring expressions whose properties are vague or debatable.

5.3.1. The role of context

As mentioned in Section 1, the work described in this paper has focused on referring expressions in isolation from their linguistic context. This has allowed us to offer precise definitions of some of the key notions involved, such as under-specification, and various types of over-specification. When the context of use is taken into account, things can become more challenging, even in short stretches of text. A miniature example is offered by Stone and Webber (1998), who discuss a situation involving two hats and two rabbits, with one of the two rabbits sitting in one of the two hats. Here, "*the rabbit in the hat*" is a distinguishing description because even though the situation involves more than one hat, the NP as a whole makes it clear which hat and which rabbit is being referred to.

When a wider stretch of context is taken into account, it is extremely common for an NP to be under-specified in isolation (i.e., when only the NP itself is taken into account) whereas, in context, it is a distinguishing description. For example, when

we say "*My father bought a dog; the dog eats sausages*", the NP "*the dog*" is under-specified in isolation, but fully specified (i.e., a distinguishing description) when the sentence as a whole is taken into account. This phenomenon has been widely discussed in the psycholinguistic (e.g., Pogue et al. (2016)) and computational literature (e.g. Krahmer and Theune (2002)).

All the concepts discussed in the present paper can, in principle, be generalised to take context into account. For example, consider our definition of a Distinguishing Description in Section 3.1: its informal part requires that such a description "singles out r from all other elements of C"; this idea remains valid when the context of the use of the description is taken into account. However, this is not true for the formal part of the definition, which requires that $[\ [P_1]\ ] \cap \ldots \cap [\ [P_n]\ ] = \{r\}$ (where r is the referent), since this narrow definition would adjudge "*the dog*" not to be a distinguishing description.

It would be interesting to generalise our definitions in such a way that contextual information can help to "disambiguate" a referring expression (i.e., to make it distinguishing). Applying these new definitions in an annotation scheme would not be without its challenges, because it would be up to the human annotators to decide whether (for example) a given NP, in a given context of use, manages to "single out" the referent. Undoubtedly, different annotators would sometimes resolve such questions differently, necessitating protocols (i.e., an annotation manual), as is often done when pragmatic information needs to be entered by human annotators[6] We discuss such "disagreements" later in this subsection.

5.3.2. Vague and debatable properties

The studies at the heart of this paper did not consider the role of debatable and vague properties in reference. All the referring expressions we discussed were constructed from content words that have — in the simple domains that were considered — clear-cut, crisp interpretations. For example, although the referring expressions in TUNA's furniture domain make use of the vague attribute SIZE, each type of furniture in the TUNA domain comes in only two sizes, with nothing in between; consequently, there was never any doubt as to what was meant by, for example, a "large" chair. Similarly, although shapes and colours in real life come in subtle variations, all the stimuli in TUNA's furniture corpus were taken from Michael Tarr's carefully curated Object Databank, where shapes and sizes are carefully controlled to be crisply distinct and easily recognisable.

Debatable properties do exist in TUNA's people domain. When we asked annotators to annotate referring expressions in the people domain, disagreements came up in relation to properties such as AGE: AGE is a vague attribute (e.g., 50 years might be on the borderline between old and not-old); moreover, the age of a person depicted in a photograph is sometimes hard to decide. Since disagreements did not occur very frequently in the people domain, we asked annotators to discuss the cases they disagreed on and to make a final decision in each case. This solution might not work if one were to apply our annotation scheme to domains that contain more debatable or vague properties.

In TUNA's people domain, the properties expressed by speakers are not only vague sometimes; they can also be debatable. An example arises when a person is referred to as "the man with the funny smile", where opinions can differ starkly on whether a smile is debatable.

Even in situations such as the ones discussed here, annotation is possible, although we could not simply rely on the definitions given in section 3. One option would be to follow the approach in Arts (2004), who asked each annotator *"Tick this box if you believe that the author could have used an alternative description, which you would have understood substantially faster. If so, then please suggest such an alternative description."* Arts (2004, Chapter 4) argued that the second rule of the Gricean Maxim of Quantity is violated by the use of a given Referring Expression if and only if the time that recipients need to identify the intended referent of that expression (i.e., identification time) is greater than the identification time associated with some alternative expression. She showed experimentally that an apparently over-specified referring expression (like "*the round button on the right*", when the situation contains only one round button) can speed up identification; she concluded that this expression does not break the second rule of Gricean Quantity. For our annotation scheme, we could ask annotators to answer questions like "*is there any property in this referring expression that is redundant?*" (for real over-specification or nominal over-specification) or "*if none of the properties is redundant, can you please suggest a shorter expression?*" (for numerical over-specification).

In this way, one could collect an annotated corpus, each referring expression in which is associated with one category (if all annotators agree) or several. When analysing the resulting corpus, we could investigate how many disagreements there are and which properties contribute to the disagreements. To aggregate disagreements between annotators, a simple option is majority voting, annotating each referring expression with the category that most annotators agree on. Another option is to model each annotation as a distribution over all possible categories (as in Castro Ferreira et al. (2016)) and to conduct an analysis based on these distributions.[7]

5.4. Multiple referents

This paper has focused on expressions that refer to one singular entity. When we refer to a *set* of entities, some additional issues come to the fore.

References to sets may be *conjunctive* or *non-conjunctive*. Conjunctive descriptions rely on a partitioning of the target set into two or more parts, which are described separately; this happens, for example, when we say "*the red chair and the blue*

[6] See e.g., the annotation scheme in Poesio and Vieira (1998).

[7] See Uma et al. (2021) for a survey of approaches to disagreements in corpus annotation.

fan"). A non-conjunctive description does not rely on such a partitioning; this happens, for example, when we say "*the (two) grey sofas*", where a set is referred to without breaking it up into its constituent parts. Conjunctive descriptions pose deep challenges to theoretical and computational linguistics (e.g., Gatt (2007)), even when only distributive references to sets are taken into account (cf., the discussion of collective references in Stone (2000)). Nonetheless, from the point of view of the present paper, with its ontology of specification types and accompanying annotation scheme, non-conjunctive references to sets of entities can be treated in much the same way as references to a single entity. Conjunctive descriptions, by contrast, give rise to new questions.

Consider the scene in Fig. 3(a), in section 3.4, focusing on description (12-a). This description can be analysed in the usual fashion, observing that the first clause ("*the red chair*") is an under-specification, whereas the second clause ("*the large left-facing table*") is a real over-specification. Now consider (12-b): is this an over-specification? Our modelling, in section 3.1, of referring expressions as *bags* of properties (i.e., where different occurrences of a given property are kept separate) suggests that it is an over-specification, because (12-b) can be simplified by "aggregating" information, yielding (12-c). Cases like this, where an opportunity for aggregation is not utilised, can be seen as another type of over-specification because they result in expressions that use one or more properties (such as *left-facing*) more often than necessary. Alternatively, one could argue that these cases do *not* involve over-specification, because (12-b) expresses the same information as (12-c), albeit in a more elaborate fashion.

(12) a. the red chair and the large left-facing table
b. the left-facing chair and the left-facing table.
c. The left-facing chair and table.

If one accepts that to miss an opportunity for aggregation is to over-specify, then what to make of *semantic* aggregation (Reape and Mellish, 1999)? Consider descriptions (13-a), (13-b), and (13-c) for example, each of which attempts to refer to two pieces of furniture. The two occurrences of *red* in (13-a) can be syntactically aggregated as in (13-b), or *semantically*, as in (13-c), because both chairs and tables are furniture, and the scene contains no other red furniture than the two target entities.

(13) a. the red chair and the red table
b. the red chair and table.
c. the red furniture

It could be argued that (13-c) is the only "minimal description" of the three, and that (13-a) and (13-b) are both over-specified, though (13-a) to a higher extent than (13-b). What is the most enlightening way to classify these phenomena, and how to encode this information in an annotation scheme, are issues for further research.

6. General conclusions

Classic theories of descriptions, from Frege (1892) and Russell (1905) onwards, have emphasised the ability of a referring expression to pick out a referent, implying a fundamental dichotomy between descriptions that are *distinguishing* (i.e., identify their referent uniquely) and ones that are not. Grice's Cooperative Principle, with its Maxim of Quantity (Grice, 1975) suggested an additional dimension, asking not only whether a description contains *enough* information to pick out a referent, but also whether a description does so with optimal efficiency (i.e., obeying both halves of Grice's Maxim of Quantity), implying an additional distinction between descriptions that are over-specified and descriptions that are not. In the present paper, we have proposed a new perspective on the different ways in which a description can manage to pick out a referent and the different ways in which it can fail to manage this. This account is more *precise* than its predecessors because it offers semantically explicit definitions of key notions such as over-specification for the first time; it is also more *fine-grained* because it distinguishes between different kinds of over-specification; finally, it is more *extensive* because it has a place for some varieties of specification (e.g., wrong specification and duplicate attribute specification) that have often been overlooked. To demonstrate how our subcategorization of all the different types of specification can engender linguistic insights, we applied it to corpora in two languages, namely the MTUNA corpus (for Mandarin) and the ETUNA corpus (for English), showing how the new scheme permits an enhanced analysis of those corpora.

We envisage several other uses of the framework, in theoretical studies of reference, in empirical studies of corpora, and in the study of differences between different cultures and demographies (see Section 1). As we have argued in section 5.3, our framework can also be used to design improved metrics for measuring the accuracy of computational models of referring, allowing researchers to make more meaningful comparisons between the output of a model and a human-produced "gold standard". In line with these thoughts, but more speculatively, we envisage that further developments of our framework, in the style of sections 5.1.1 and 5.1.2, could make a modest but necessary contribution to the evaluation of Generative AI systems (such as the ones in the ChatGPT tradition of Aydın and Karaarslan (2022) for example), helping researchers to assess what a given system "gets right" and what it does not.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

We are grateful to the reviewers of the *Journal of Pragmatics* for their insightful remarks. We would also like to thank Ruud Koolen for his very helpful suggestions.

A Appendix. An Annotation Scheme that Utilises our Formal Account

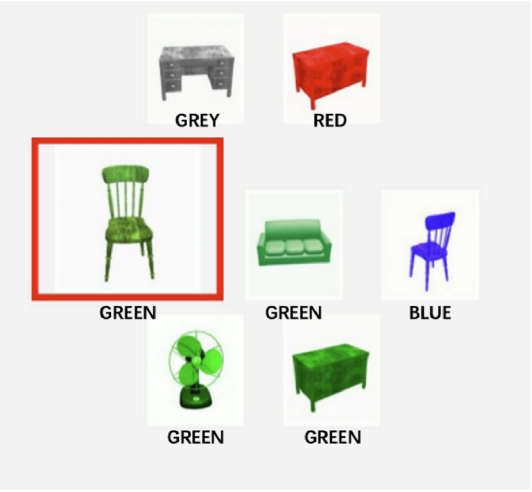


Fig. 7A scene from the MTUNA corpus.

We used the perspective that was presented formally in Section 3 as the basis for an annotation scheme for corpora, such as the TUNA corpora, where the simplifying assumptions that were made in the Introduction hold. For example, the scheme does not take differences in contextual salience between objects into account (Section 5.3.1). In this Appendix, we illustrate the scheme, using the trial in figure 7 as an example.

Each description is annotated using a set of key–value pairs. The following is a simple example of the annotation of a real over-specification "*the large green one*" in which "*green*" is superfluous.

```
LABEL: Real Over-specification

SUPERFLUOUS: 1
```

LABEL indicates what type of specification the current description belongs to, and SUPERFLUOUS records the number of superfluous properties. Below, we illustrate our annotation scheme for more complex situations.

Numerical over-specification. A numerical over-specification for figure 7 is "*the green chair*". Because none of the properties in this numerically overspecified description is superfluous (Section 3.3.1), the value of SUPERFLUOUS is set to zero.

```
LABEL: Numerical Over-specification

SUPERFLUOUS: 0
```

Superfluous TYPE. Both nominal over-specification and real over-specification can have a superfluous TYPE. However, in many analyses, we need to track the number of these superfluous TYPEs. To track them and to differentiate them from other over-specified properties, we employed a new variable, namely, SUPERFLUOUS-TYPE. If a superfluous TYPE is found, we increased the value of SUPERFLUOUS-TYPE by one. For example, for the nominal over-specification *"the large chair"*, we have:

```
LABEL: Nominal Over-specification

SUPERFLUOUS: 0

SUPERFLUOUS-TYPE: 1
```

Duplicate-attribute over-specification. To count the number of duplicate properties, we designed a new variable: DUPLICATE. For example, suppose we have three properties: $P_i, P_j, P_k \in \mathcal{D}$ (each of which could be either a TYPE or other types of property) and $P_i = P_j = P_k$, then we annotate: DUPLICATE : 2.

The number of superfluous properties. When deciding the number of superfluous properties, we counted the maximum number of properties (including TYPE) that can be removed while the resulting description is still a distinguishing description. For example, for the scene in Fig. 7, we could say:

(14) a. the front-facing green chair
b. the large green chair

From description (14-a), only the phrase *front facing* can be removed, so there is 1 superfluous property. Removing superfluous properties will sometimes result in a numerical over-specification. As for the description (14-b), either removing the *large* or removing both the *green* and *chair* yields distinguishing descriptions. However, based on the principle above, the latter removal removes more properties than the former one.

Description Basis. Following on from the previous example, to record that the choice results in a minimal description rather than a numerical over-specification, we make use of the idea of "description basis". In our annotation, we used a variable BASIS to track which type of specification (minimal description or numerical over-specification) the current description is "built around".

Here, we provide an example which contains (a real over-specification contains a superfluous TYPE property or duplicated properties). For example, suppose we have the following descriptions:

(15) the backward large table whose drawer is not visible

Compared to the minimal description "*the large one*", the over-specification (15) is a real over-specification, in which there is a superfluous TYPE (large), and two superfluous ORIENTATION (backward and whose drawer is not visible). Interestingly, the duplicated properties itself is a superfluous property. In such a case, we add SUPERFLUOUS with one for acknowledging the superfluous ORIENTATION. It is also a duplicate-attribute over-specification. We add DUPLICATE with one for its duplicated use of ORIENTATION. Therefore, the annotation of description (15) is:

```
LABEL: Real Over-specification, Duplicate-attribute Over-specification

SUPERFLUOUS: 1

SUPERFLUOUS-TYPE: 1

DUPLICATE: 1

BASIS: Minimal Description
```

Under-specification. As for the under-specifications, we used a variable named UNDERSPECIFIED to record the number of under-specified properties. For instance, for the pure under-specification "the chair", we have UNDERSPECIFIED: 1.

When deciding the amount of under-specified properties, we asked how many properties would minimally have to be added to make the description distinguishing. To do so, suppose there are two possible fixes, both of which add the same number of properties. We chose the one that generates superfluous properties as little as possible and recorded that number. For example, we can make "*the chair*" distinguishing by either adding *large* or *green*. But by adding *large*, the fixed description is actually a nominal over-specification with a superfluous TYPE. In contrast, by adding *green*, the resulting description is a numerical over-specification without any superfluous properties. We, therefore, choose the latter option and mark the description as a "Pure Under-specification" that is based on the "Numerical Over-specification":

```
LABEL: Pure Under-specification

UNDERSPECIFIED: 1

BASIS: Numerical Over-specification
```

Meanwhile, because of the existence of mixed descriptions, we also record the number of superfluous properties.

In addition, we argue that, for under-specifications, it is uninteresting to record whether a superfluous property is a TYPE or not and whether it is a duplicated property or not. Therefore, we used only one variable (i.e., "SUPERFLUOUS") to track the number of superfluous properties in under-specifications.

References

Arts, A., 2004. Overspecification in Instructive Texts. Wolf, Nijmegen.
Arts, A., Maes, A., Noordman, L., Jansen, C., 2011. Overspecification facilitates object identification. J. Pragmat. 43 (1), 361–374.
Aydın, Ö., Karaarslan, E., 2022. Openai Chatgpt Generated Literature Review: Digital Twin in Healthcare. Available at SSRN 4308687.
Belz, A., Kow, E., Viethen, J., Gatt, A., 2010. Generating referring expressions in context: the GREC task evaluation challenges. In: Krahmer, E., Theune, M. (Eds.), Empirical Methods in Natural Language Generation: Data-Oriented Methods and Empirical Evaluation, Volume 5790 of Lecture Notes in Computer Science. Springer, pp. 294–327.

Castro Ferreira, T., Krahmer, E., Wubben, S., 2016. Towards more variation in text generation: developing and evaluating variation models for choice of referential form. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Berlin, Germany, pp. 568–577.
Chao, Y.R., 1965. A Grammar of Spoken Chinese. Univ of California Press.
Chen, G., 2022. Computational Generation of Chinese Noun Phrases. PhD thesis. Utrecht University.
Clark, H.H., 1996. Using Language. Cambridge University Press.
Clark, H.H., Murphy, G.L., 1982. Audience design in meaning and reference. In: Advances in Psychology, vol. 9. Elsevier, pp. 287–299.
Dale, R., 1989. Cooking up referring expressions. In: 27th Annual Meeting of the Association for Computational Linguistics, pp. 68–75.
Dale, R., 1992. Generating Referring Expressions: Building Descriptions in a Domain of Objects and Processes.
Dale, R., Reiter, E., 1995. Computational interpretations of the gricean maxims in the generation of referring expressions. Cognit. Sci. 19 (2), 233–263.
Dale, R., Viethen, J., 2009. Referring expression generation through attribute-based heuristics. In: Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009), pp. 58–65.
Deemter, K.v., Gatt, A., Sluis, I.v. d., Power, R., 2012. Generation of referring expressions: assessing the incremental algorithm. Cognit. Sci. 36 (5), 799–836.
Degen, J., Hawkins, R.D., Graf, C., Kreiss, E., Goodman, N.D., 2020. When redundancy is useful: a bayesian approach to "overinformative" referring expressions. Psychol. Rev. 27 (4), 591–621.
Dice, L.R., 1945. Measures of the amount of ecologic association between species. Ecology 26 (3), 297–302.
Eikmeyer, H.-J., Ahlsén, E., 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, P.E., Bailey, K.G., Ferreira, F., 2006. Do speakers and listeners observe the Gricean maxim of quantity? J. Mem. Lang. 54 (4), 554–573.
Engelhardt, P.E., Demiral, Ş.B., Ferreira, F., 2011. Over-specified referring expressions impair comprehension: an ERP study. Brain Cognit. 77 (2), 304–314.
Ferreira, V.S., Slevc, L.R., Rogers, E.S., 2005. How do speakers avoid ambiguous linguistic expressions? Cognition 96 (3), 263–284.
Ford, W., Olson, D., 1975. The elaboration of the noun phrase in children's description of objects. J. Exp. Child Psychol. 19 (3), 371–382.
Frank, M.C., Goodman, N.D., 2012. Predicting pragmatic reasoning in language games. Science 336 (6084), 998, 998.
Frege, G., 1892. Ueber sinn und bedeutung. Zeitschrift für Philosophie und philosophische Kritik 100 (3), 25–50.
Fukumura, K., van Gompel, R.P., 2011. The effect of animacy on the choice of referring expression. Lang. Cognit. Process. 26 (10), 1472–1504.
Garoufi, K., Koller, A., 2014. Generation of effective referring expressions in situated context. Language, Cognition and Neuroscience 29 (8), 986–1001.
Gatt, A., 2007. Generating Coherent References to Multiple Entities. PhD thesis. Citeseer.
Gatt, A., Belz, A., 2009. Introducing shared tasks to nlg: the tuna shared task evaluation challenges. In: Empirical Methods in Natural Language Generation. Springer, pp. 264–293.
Gatt, A., van der Sluis, I., van Deemter, K., 2007. Evaluating algorithms for the generation of referring expressions using a balanced corpus. In: Proceedings of the Eleventh European Workshop on Natural Language Generation (ENLG 07). DFKI GmbH, Saarbrücken, Germany, pp. 49–56.
Goodman, N.D., Frank, M.C., 2016. Pragmatic language interpretation as probabilistic inference. Trends Cognit. Sci. 20 (11), 818–829.
Grice, H.P., 1975. Logic and Conversation, pp. 41–58, 1975.
Gupta, S., Stent, A., 2005. Automatic evaluation of referring expression generation using corpora. In: Proceedings of the Workshop on Using Corpora for Natural Language Generation. Citeseer, pp. 1–6.
Horton, W., Keysar, B., 1996a. When do speakers take into account common ground? Cognition 59, 91–117.
Horton, W.S., Keysar, B., 1996b. When do speakers take into account common ground? Cognition 59 (1), 91–117.
Howcroft, D.M., Vogels, J., Demberg, V., 2017. G-tuna: a corpus of referring expressions in German, including duration information. In: Proceedings of the 10th International Conference on Natural Language Generation, pp. 149–153.
Huang, C.-T.J., 1984. On the distribution and reference of empty pronouns. Linguistic Inquiry 15 (4), 531–574.
Jara-Ettinger, J., Rubio-Fernandez, P., 2021. The social basis of referential communication: speakers construct physical reference based on listeners' expected visual search. Psychol. Rev.
Jordan, P.W., Walker, M.A., 2005. Learning content selection rules for generating object descriptions in dialogue. J. Artif. Intell. Res. 24, 157–194.
Keysar, B., Lin, S., Barr, D.J., 2003. Limits on theory of mind use in adults. Cognition 89 (1), 25–41.
Kibrik, A.A., Khudyakova, M.V., Dobrov, G.B., Linnik, A., Zalmanov, D.A., 2016. Referential choice: predictability and its limits. Front. Psychol. 7, 1429.
Koolen, R., Gatt, A., Goudbeek, M., Krahmer, E., 2011. Factors causing overspecification in definite descriptions. J. Pragmat. 43 (13), 3231–3250.
Koolen, R., Goudbeek, M., Krahmer, E., 2013. The effect of scene variation on the redundant use of color in definite reference. Cognit. Sci. 37 (2), 395–411.
Koolen, R., Krahmer, E., 2010. The D-Tuna Corpus: A Dutch Dataset for the Evaluation of Referring Expression Generation Algorithms. LREC.
Krahmer, E., Theune, M., 2002. Efficient context-sensitive generation of referring expressions. Information Sharing: Reference and Presupposition in Language Generation and Interpretation 143, 223–263.
Krahmer, E., van Deemter, K., 2012. Computational generation of referring expressions: a survey. Comput. Ling. 38 (1), 173–218.
Kutlak, R., Van Deemter, K., Mellish, C., 2016. Production of referring expressions for an unknown audience: a computational model of communal common ground. Front. Psychol. 7, 1275.
Levelt, W.J., 1993. Speaking: from Intention to Articulation, vol. 1. MIT press.
Lv, S., 1979. Problems in the Analysis of Chinese Grammar.
Matthews, D., Lieven, E., Theakston, A., Tomasello, M., 2006. The effect of perceptual availability and prior discourse on young children's use of referring expressions. Appl. Psycholinguist. 27 (3), 403–422.
Matthews, D., Lieven, E., Theakston, A., Tomasello, M., 2016. The effect of perceptual availability and prior discourse on young children's use of referring expressions. Appl. Psycholinguist. 27 (3), 403–422.
McDonald, D.D., 1983. Description directed control: its implications for natural language generation. Comput. Math. Appl. 9 (1), 111–129.
Mill, J.S., 1843. A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence, and the Methods of Scientific Investigation. John W. Parker, London.
Monroe, W., Potts, C., 2015. Learning in the rational speech acts model. arXiv preprint arXiv:1510.06807.
Neale, S.R.A., 1988. Descriptions. Stanford University.
Newnham, R., 1971. About Chinese. Penguin Books Ltd.
Olson, D.R., 1970. Language and thought: aspects of a cognitive theory of semantics. Psychol. Rev. 77 (4), 257.
Paraboni, I., Lan, A.G.J., de Sant'Ana, M.M., Coutinho, F.L., 2017. Effects of cognitive effort on the resolution of overspecified descriptions. Comput. Ling. 43 (2), 451–459.
Paraboni, I., van Deemter, K., 2014. Reference and the facilitation of search in spatial domains. Language, Cognition and Neuroscience 29 (8), 1002–1017.
Pechmann, T., 1989. Incremental speech production and referential overspecification. Linguistics 27 (1), 89–110.
Poesio, M., Vieira, R., 1998. A corpus-based investigation of definite description use. Comput. Ling. 24 (2), 183–216.
Pogue, A., Kurumada, C., Tanenhaus, M.K., 2016. Talker-specific generalization of pragmatic inferences based on under-and over-informative prenominal adjective use. Front. Psychol. 6, 2035.
Reape, M., Mellish, C., 1999. Just what is aggregation anyway. In: Proceedings of the 7th European Workshop on Natural Language Generation. Citeseer, pp. 20–29.
Rehrig, G., Cullimore, R.A., Henderson, J.M., Ferreira, F., 2021. When more is more: redundant modifiers can facilitate visual search. Cognitive Research: Principles and Implications 6 (1), 1–20.
Reiter, E., 2018. A structured review of the validity of bleu. Comput. Ling. 44 (3), 393–401.
Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., Boyes-Braem, P., 1976. Basic objects in natural categories. Cognit. Psychol. 8 (3), 382–439.

Rubio-Fernandez, P., 2021. Color discriminability makes over-specification efficient: theoretical analysis and empirical evidence. Humanities and Social Sciences Communications 8 (1), 1–15.
Russell, B., 1905. On denoting. Mind 14 (56), 479–493.
Schriefers, H., Pechmann, T., 1988. Incremental production of referential noun phrases by human speakers. Advances in Natural Language Generation 1, 172–179.
Searle, J., 1969. Speech Acts: an Essay in the Philosophy of Language. Cambridge University Press, Cambridge, UK.
Sonnenschein, S., 1984. The effects of redundant communications on listeners: why different types may have different effects. J. Psycholinguist. Res. 13 (2), 147–166.
Stone, M., 2000. On identifying sets. In: INLG'2000 Proceedings of the First International Conference on Natural Language Generation, pp. 116–123.
Stone, M., Webber, B., 1998. Textual economy through close coupling of syntax and semantics. arXiv preprint cmp-lg/9806020.
Strawson, P.F., 1950. On referring. Mind 59 (235), 320–344.
Tourtouri, E.N., Delogu, F., Sikos, L., Crocker, M.W., 2019. Rational over-specification in visually-situated comprehension and production. Journal of Cultural Cognitive Science 3 (2), 175–202.
Uma, A.N., Fornaciari, T., Hovy, D., Paun, S., Plank, B., Poesio, M., 2021. Learning from disagreement: a survey. J. Artif. Intell. Res. 72, 1385–1470.
van Deemter, K., 2016. Computational Models of Referring: a Study in Cognitive Science. MIT Press.
van Deemter, K., Sun, L., Sybesma, R., Li, X., Chen, B., Yang, M., 2017. Investigating the content and form of referring expressions in Mandarin: introducing the MTUNA corpus. In: Proceedings of the 10th International Conference on Natural Language Generation. Association for Computational Linguistics, Spain, pp. 213–217. Santiago de Compostela.
van der Sluis, I., Krahmer, E., 2007. Generating multimodal references. Discourse Process 44 (3), 145–174.
van Gompel, R.P., van Deemter, K., Gatt, A., Snoeren, R., Krahmer, E.J., 2019. Conceptualization in reference production: probabilistic modeling and experimental testing. Psychol. Rev. 126 (3), 345.
Villalba, M., Teichmann, C., Koller, A., 2017. Generating contrastive referring expressions. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Association for Computational Linguistics, Vancouver, Canada, pp. 678–687.
Yu, L., Poirson, P., Yang, S., Berg, A.C., Berg, T.L., 2016. Modeling context in referring expressions. In: Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer, pp. 69–85.
Yu, L., Tan, H., Bansal, M., Berg, T.L., 2017. A joint speaker-listener-reinforcer model for referring expressions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7282–7290.

Guanyi Chen Guanyi Chen holds the position of Associate Professor at the School of Computer Science at Central China Normal University, Wuhan, China. He was a lecturer at the Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands. He received his PhD also from Utrecht University and received an M.S. degree in artificial intelligence from the University of Edinburgh, Edinburgh, U.K., in 2016. His research interests mainly lie in natural language generation.

Kees van Deemter Kees van Deemter leads the Natural Language Processing group at Utrecht University's Computing and Information Sciences department. He has led a number of projects in Natural Language Generation, both in industry (at Philips Research from 1984 till 1996) and in academia (at the University of Brighton and at the University of Aberdeen). Working with linguists, psychologists and logicians, he has published widely on theoretical issues in his field, including the monographs Not Exactly: in Praise of Vagueness (Oxford University Press 2010), and Computational Models of Referring: a Study in Cognitive Science (MIT Press 2016).