Linguistic **I**ssues in **L**anguage **T**echnology – **LiLT**
Submitted, June 2011

Issues in Translating and Producing Japanese Referring Expressions for Dialogues

Ielka van der Sluis
Saturnino Luz

Published by CSLI Publications

LiLT volume 5, issue 1 June 2011

Issues in Translating and Producing Japanese Referring Expressions for Dialogues

Ielka van der Sluis, *Trinity College Dublin,* Saturnino Luz, *Trinity College Dublin*

Abstract

This paper presents an analysis of cross-cultural and linguistic issues in the generation of referring expressions (REs) for English and Japanese dialogues. The analysis is based on a translation activity carried out for a study on the perception of automatically generated REs in a virtual world involving participants in Tokyo and Dublin and complemented by two data elicitation studies. In order to preserve the output of the RE generation algorithm the translation sought to produce a Japanese dialogue in which the REs were as close to the English originals as possible, but within a scenario adapted to the Japanese culture. The two data elicitation experiments assessed Japanese speakers' preferences of REs in the same dialogue context. Insights from this work are relevant to the design of RE generation algorithms for use in real-life situations and raise questions as to what extent the current algorithms are transferable between languages and cultures. Results suggest that current approaches to generating REs are biased towards the English language, which becomes apparent when considering, for instance, the realisation of Japanese locative expressions and the absence of a distinction between singular and plural in the Japanese language.

LiLT Volume 5, Issue 1, June 2011.
Issues in Translating and Producing Japanese Referring Expressions for Dialogues.

1 Introduction

1.1 Background

The generation of referring expressions (GRE) is a central task in Natural Language Generation (NLG), and various algorithms which automatically produce referring expressions (RE) have been developed. Recent examples include (Gardent, 2002, Van Deemter, 2002, Krahmer et al., 2003, Jordan and Walker, 2005, Funakoshi et al., 2006, Van Deemter, 2006). Existing GRE algorithms generally assume that both speaker and addressee have access to the same information. In most cases this information is represented by a knowledge base that contains the objects and their properties which are present in the domain of conversation in terms of attribute-value pairs. A typical algorithm takes as input a single object (the target) and a set of objects (the distractors) from which the target object needs to be distinguished (cf. Dale and Reiter, 1995). The task of a GRE algorithm is to determine which set of properties is needed to single out the target from the distractors.

Many of these algorithms focus on 'first-mention' REs referring to objects that have not been talked about earlier in the discourse. However, some algorithms have addressed the way in which the utterance context affects the choice of REs (e.g., the work on the type of noun phrase suitable in a given context by McCoy and Strube, 1999, Henschel et al., 2000, Jordan, 2000, Callaway and Lester, 2002, Poesio et al., 2004, Byron and Stoia, 2005, Gardent and Striegnitz, 2007, Janarthanam and Lemon, 2010)). The generation of full noun phrases in a linguistic context has also highlighted important factors such as salience and centering (Krahmer and Theune, 2002, Siddharthan and Copestake, 2004, Gupta and Stent, 2005, Jordan and Walker, 2005). Nevertheless many issues remain, particularly in relation to reference in situated dialogue, where the interaction between speakers turns reference into a joint enterprise in which speakers influence and complement each other in their perception of the conversation domain.

Since human communication includes gestures as well as language some GRE research has also focussed on REs that include pointing gestures, and various algorithms for the generation of such multimodal REs have been proposed (cf. André and Rist, 1993, Claassen, 1992, Kranstedt et al., 2006, Lester et al., 1997, Reithinger, 1992). In the study described in this paper we take the algorithm by Van der Sluis and Krahmer (2007), a multimodal variant of the algorithm proposed by (Krahmer et al., 2003), as a starting point. This multimodal algorithm approaches GRE as a compositional task in which language and gestures are combined in a flexible way to identify objects.

Over the last decade, evaluation of these GRE algorithms has become more important as can be observed through the success of various shared tasks (Gatt and Belz, 2010, Belz and Kow, 2010). Evaluation of GRE algorithms traditionally use intrinsic methods to evaluate automatically generated output against human produced output as collected in corpora (cf. Belz, 1994). In addition, evaluation of multimodal GRE algorithms has commonly made use of data elicited in settings in which people were instructed to use pointing gestures while identifying objects (Van der Sluis and Krahmer, 2004, Kranstedt et al., 2006). Although there has been increased interest in extrinsic evaluation methods which consider the effects of the output on something external, such as human judgement (Belz and Gatt, 2008, Belz and Reiter, 2009, Belz et al., 2010), current extrinsic evaluations are limited to artificial settings in which humans are asked to perform tasks outside their daily practice. We believe that in settings which involve complex contextual elements, such as generation across languages and cultures, evaluation may be best tackled by a combination of methods. In this paper we combine elements of extrinsic evaluation, qualitative methods and intrinsic measures to assess the extent to which a GRE algorithm is transferable between languages and cultures, in this case English and Japanese. The paper therefore starts with a qualitative analysis which presents a whole range of issues regarding the realisation of REs in Japanese, from a translation perspective, and then focuses on an evaluation of first-mention multimodal REs in a dialogue context in which the dialogue partners physically take part in the domain of conversation.

1.2 Context and Outline of Work in this Paper

The work presented in this paper is part of a cross-cultural investigation on human perception of automatically generated multimodal REs in a virtual world. More details about this project in its broader scope is given elsewhere (Breitfuss et al., 2009, Van der Sluis et al., To appear). In the present paper, we focus on the analysis of cross-cultural and linguistic issues for generating the REs that appeared in the dialogue used in the project. The dialogue was originally written in English and subsequently evaluated and translated into Japanese. The goal of the translation was to produce a Japanese dialogue in which the REs were as close to the English originals as possible in order to preserve the output of the GRE algorithm. However, the dialogue scenario itself was adapted so as to adhere to cultural norms and perceptions of a Japanese context, thereby minimising the effects that extraneous variables (i.e. variables other than the choice of RE generation strategy) might have on the results of the project's perception study.

The translation process revealed various issues about the use of REs which could impact on the participants' perceptions of the output of GRE algorithms. These issues related to the utility of attributes for object identification, the realisation of locative expressions, and the absence of a singular/plural distinction in Japanese. Our aim in this paper, is to present a detailed analysis of these issues, and discuss their implications to the design of GRE algorithms for use in interactive applications. In particular, we discuss the transferability of current algorithms across languages, and suggest that the evaluation of GRE output is more complicated than currently thought. Computational work on GRE has heavily used similarity metrics for measuring the 'quality' of the output of algorithms (see also the shared tasks and evaluation challenges in this area). Such evaluations are based on a comparison of semantic representations of automatically produced REs and human produced REs. These semantic representations abstract away from syntactic structures and lexical realisations (Van Deemter and Gatt, 2009). As we will see in Section 3), syntactic structure and realisation have a considerable impact on the perception of REs.

Our initial qualitative analysis of the translation of the REs was followed by two data elicitation studies (i.e., Study I and Study II) which assessed Japanese speakers' preferences of the REs used in the aforementioned dialogue. Study I was set up as a focus group study, and Study II as a web based experiment. Participants in both elicitation studies were asked to compose REs at particular points in the dialogue to indicate objects in a domain that was presented to them in a schematic way. The REs, could be composed by selection of a determiner or demonstrative and selection of a linguistic description from a list of available options. Results show unexpected differences the use of demonstratives and, most notably, a preference for linguistic descriptions that do not uniquely identify the target. This is remarkable compared to elicitation studies in the English language (cf. Viethen and Dale, 2006), which show that people commonly include more properties than strictly necessary to distinguish a target object.

The paper is structured as follows: Firstly, the next section gives a description of the multimodal GRE algorithm and the dialogue used in the study, accompanied by the preliminary comments and requests for clarification made by the Japanese translator in preparation for the task. Secondly, we present an analysis of the issues which arose in the process of translating REs from English into Japanese. Thirdly we present the two data elicitation studies. And finally, the paper closes with a discussion of the implications of this work for research in GRE.

2 Method and setting

2.1 An Algorithm for Generating Multimodal REs

The multimodal GRE algorithm by Van der Sluis and Krahmer (2007) which was taken as a starting point for this study approaches GRE as a compositional task in which language and gestures can be combined in a flexible way to identify a target. Following Krahmer et al. (2003), the algorithm generates a RE by searching through a graph-based representation of the domain of conversation for the best sub-graph that uniquely identifies the target object, where 'best' refers to the effort involved in verbal production and pointing. In the graph-based representation of the domain, the objects are represented as vertices and the properties and relations of the objects are represented as edges. For each RE to be generated the domain graph is enriched with a gesture graph, which includes edges for pointing gestures directed to the target object. The scope of a pointing gesture depends on the distance between the target object and the pointing device (in this paper, this would be the finger or hand of a virtual agent). For instance, if the target object is near to the agent, it may be uniquely identified by a pointing gesture. However, the larger the distance between the target and the agent, the more distractor objects may be included in the scope of the pointing gesture.

Consequently, in a virtual world, the algorithm can cause an agent to identify an object located far away by moving closer to the object so as to distinguish it with a very 'precise' (i.e. uniquely identifying) pointing gesture and the use of limited linguistic information (e.g., 'this one'). Alternatively, the algorithm could generate an 'imprecise', pointing gesture including other objects in its scope. In this case, more linguistic information has to be added to the RE to ensure that the object can be uniquely identified by the addressee. A virtual character could say, for instance, 'the large blue desk in the back' and accompany this description with an 'imprecise' pointing gesture towards a desk surrounded by other objects and located at a distance from the agent, where the imprecise gesture still serves to decrease the number of distractor objects.

The algorithm determines the multimodal content of the RE, by searching for the best (i.e. least costly) distinguishing graph that identifies the target object. It uses a cost function to determine which linguistic properties and/or pointing gestures to include based on a notion of effort. The cost of the linguistic properties can be determined, for instance, on the basis of human preferences, as in the use of the list of preferred properties in the incremental algorithm proposed by Dale and Reiter (1995). Alternatively, a search for the shortest RE can be mimicked by assigning all properties the same cost. The cost of the pointing gestures is determined by Fitts' Law (Fitts, 1954) employing the size of the target (e.g., large objects are generally easier to point out than small objects) and the distance between the pointing device and the target (e.g., objects that are located at a close distance are generally easier to point out than far away objects). In this paper we will focus on pointing gestures performed by stationary agents. Although some of the data presented in Section 4 may be used to inform relations between linguistic properties and pointing gestures in terms of costs, we will not pursue this issue but concentrate on the linguistic output. For a more detailed description of the algorithm we refer to the paber by Van der Sluis and Krahmer (2007).

2.2 Dialogue and Setting

We employed a Fully Generated Scripted Dialogue (FGSD) approach (André et al., 2000, Rist et al., 2003, Williams et al., 2007) to evaluate the output of the Van der Sluis & Krahmer algorithm. With FGSD entire dialogues are produced by one generator. Initially, scripted dialogues made heavy use of canned text, but recently this approach has been integrated with Natural Language Generation techniques (Van Deemter et al., 2008, Piwek, 2008). FGSD allows us to produce dialogues, without implementing a full natural language interpretation module. For our study a dialogue script was written by hand for two virtual agents in a virtual furniture store. The virtual furniture shop contains over 40 objects of which some were used as target referents in the dialogue and others were used as distractor objects. The furniture domain was chosen because detailed data on how humans refer to furniture is available through the COCONUT corpus (Di Eugenio et al., 2000) and the TUNA corpus (Van Deemter et al., To Appear), and we hoped that these knowledge sources would help us to construct a believable dialogue. The dialogue consists of 19 utterances and features a conversation between a female agent purchasing furniture for her office, and a male shop-owner guiding her through the store while describing some furniture items. Results from a pilot study used for validation of the dialogue and the setting showed that the dialogue was acceptable to an English speaking audience.

The virtual setting of the dialogue enabled us to choose a specific domain of conversation in which all objects and their properties are known. This allows for complete semantic and pragmatic transparency, which is important for a content determination task like the generation of REs. The dialogue was used as a template in which five first-mention REs were varied. The multimodal REs used to fill out these slots can

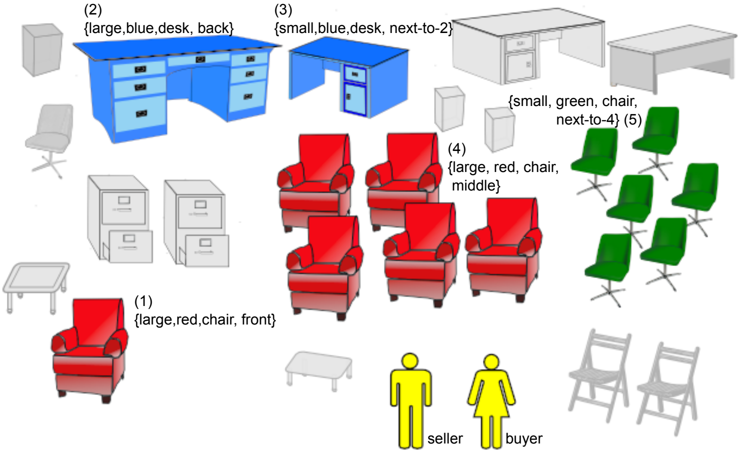


Figure 1 Bird's-eye sketch of the virtual furniture shop.

be automatically generated with the Van der Sluis & Kramer algorithm and were chosen very carefully to cover various aspects of REs as are currently being studied: (1) cardinality, the REs targeted three singular objects and two larger sets of items; (2) locative expressions, the REs included three absolute locative expressions and two relative locative expressions; and (3) the position of the referent, the target referents were distributed in the domain of conversation such that one target was located near, and two targets were located far away from the agents, and two sets of targets were located somewhere in between those two extremes with respect to the initial position of agents (i.e., in absolute locative terms from the perspective of the furniture seller agent, the singular red chair was located in the front, the two sets of chairs were located in the middle and the two desks were located in the back of the shop).

Figure 1 presents a schematic layout of the virtual furniture shop marking the positions of the agents and the furniture items. The figure shows 14 furniture items that are used for assessing multimodal GRE output: (1) a large red chair (bottom left); (2) a large blue desk (top left), (3) a small blue desk (next to the large one); (4) a set of five large red chairs (in the middle), and (5) a set of six small green chairs (next to the set of red ones), as well as a number of distractors (greyed-out items). The specific realisations of REs for the objects listed above will be henceforth referred to as *RE1*, . . . , *RE5*, respectively. The dialogue script was implemented so that the furniture seller produces linguistic descriptions in combination with deictic gestures that point in the direction of the target objects. It is assumed that the agents stay stationary at the position indicated in Figure 1 and point in the direction where the target item(s) can be found. The translation of the dialogue discussed in Section 3 considers REs which contain all object attributes known to the algorithm. The remainder of this paper only addresses the linguistic parts of the multimodal REs shown in Table 1.

2.3 Preliminary Observations Concerning the Setting

The Japanese language, especially when used in dialogue, is extremely dependent on the relationship between the dialogue partners, their gender, age and social standing. Likewise, it is expected that the appearance of the virtual furniture shop, its standing and size will influence the language and the attitudes of the people in it. This is, of course, also true of English though perhaps not to the same extent as Japanese. The importance of such factors became evident to us during the initial stage of the study, when the translator observed that "[the virtual agents had] a sort of abstract countenance that seems to belong to anywhere and thus nowhere". According to her, this posed a problem for the translation, because the language a character uses is part of its personality and by cohering a character's personality and behaviour the sense of 'presence' or 'life' that can be felt by the audience is strengthened. Therefore, despite the fact that the setting was positively assessed in a pilot study conducted with English speakers, we decided to use more localised agents in the Japanese as well as in the English set up, and to make the agents look between 25 to 35 years old. The relationship between the agents was defined as a 'shop owner - office lady' relationship and the kind of furniture shop as an average middle-end shop. All this implied that the kind of Japanese language used by the agents should be a socially polite form whose modern usage, especially among the younger generation of Japan, does not include much difference between the masculine and feminine form. For further details on the virtual world, the graphical representations, a detailed description of the environment including the way the agents could move through it can be found see Van der Sluis et al. (To appear).

3 From English to Japanese

3.1 Translation and Localisation

The analysis of translation issues presented in this section is the result of a close collaboration with a professional Japanese translator over a period of over two months. The process focused on the text of the dialogue and contextualised with the help of our animated 3D scene created in a virtual environment. The goal of the translation was to produce a Japanese dialogue in which the REs were as close to the English originals as possible. The scenario, however, in which a lady enters a shop looking for furniture was adapted to the Japanese style such that the Japanese audience could easily conceive the situation. For this

TABLE 1 Referring expressions in English, their Japanese translations, the phonetic descriptions of the Japanese translations, an indication of word order of attributes, and a retranslation to English.

RE1	*the large red one in the front* (where, ‘one’ = chair) こちら の 手前 の 大きな 赤い イス kochira no temae no ookina akai isu front large red chair Retranslation: large red chair in near direction/place in front.
RE2	*the large blue desk in the back* あちら の 奥 の 大きな 青い 机 achira no oku no ookina aoi tsukue back large blue desk Retranslation: large blue desk in far direction/place in back
RE3	*the small blue desk next to it* (where ,‘it’ = ‘the large blue desk in the back’, ie. the object referred to by 2) その （大きな青い机の） 隣 の 小さな 青い 机 sono (ookina aoi tsukue no) tonari no chiisana aoi tsukue (omitted: large blue desk) next small blue desk Retranslation: small blue desk next to none other than (large blue desk)
RE4	*the large red chairs in the middle* そちら の 中程 の 大きな 赤い イス sochira no nakahodo no ookina akai isu middle large red chair(s) Retranslation: large red chair/chairs in not too far or too near place/direction in middle
RE5	*the small green chairs next to the red ones* (where, ‘ones’ = ‘the large red chairs in the middle’, ie. the objects referred to by 4) 赤い イス の 隣 の 小さな 緑色 の イス akai isu no tonari no chiisana midori-iro no isu red chair(s) next small green-colour chair(s) Retranslation: small green-colour chair/chairs next to red chair/chairs

reason, the beginning of the dialogue was altered considerably in the Japanese version, both in speech and in gestures. For instance, in the Dublin version the furniture seller opens the dialogue with ‘Hi, how can I help you?’ accompanied with no particular gesture. In the Japanese version, the furniture seller says: ‘Irasshai-mase’ meaning ‘Welcome to our shop’ accompanied with a bow of 30 degrees. A literal translation of ‘how can I help you’, was considered to be too assertive. In general, the English dialogue was considered verbose if compared to a typical Japanese equivalent. The Japanese language (especially colloquial language) has a great tendency to omit, abbreviate and to positively use ‘silence’, or in other words, to trust in the addressee’s ability to comprehend the implications of the unspoken words. In what follows the use of pronouns and anaphora, the choice of attributes, the realisation of *location* and cardinality is addressed in more detail.

3.2 Pronouns and Anaphora

An English RE can be realised using ‘one’ instead of a full head noun, N, when the context of a description contains another NP whose head is also N (cf. Dale, 1992). Pronouns like ‘it’ can be also generated, depending on the salience of the target object in its context (Krahmer and Theune, 2002). However, in translating the REs from English to Japanese, a number of issues arose, such as the fact that anaphoric expressions like ‘one’ or ‘the ones’, which were used to indicate the set of red chairs in ‘the small green chairs next to the red ones’ and the pronouns ‘it’, which was used to indicate the large blue desk in ‘the small blue desk next to it’, do not have a precise equivalent in Japanese. Although the Japanese word ‘mono’ is often used to replace English words ‘one’ or ‘ones’, and the Japanese ‘sore’ often replaces the English ‘it’, the functions of these Japanese words differ from those of their English counterparts. Furthermore, in practice translators often omit ‘mono’ or ‘sore’ to produce a naturally flowing text. Alternatively, ‘one’ or ‘it’ may also be translated to a non-pronominal RE to the antecedent. See also the work on centering in Japanese (e.g. Walker et al., 1994, Kameyama, 1997, Iida, 1997).

Table 1 shows that ‘one’, in expression RE1, was translated as ‘isu’ meaning ‘chair’. The pronoun ‘it’ in expression RE3, has been omitted. In this expression, ‘it’ refers to ‘the large blue desk in the back’ which was included in the preceding utterance. Thus, the referent of ‘it’ is already implied by the text. This is emphasised by the use of ‘sono’, which in this case means ‘none other than [what was mentioned earlier]’. In expression RE5, the NP ‘the red ones’ was not omitted because it includes not only implicit information about the type of the referents, but also about the colour and cardinality of the referent. In this case, the red colour contrasts with the green colour of the target objects of the RE (e.g. the small green chairs). In addition, because Japanese lacks the morphological means to indicate plurality, the translator sought to retain at least the colour information. In general, our translator chose translations that felt most natural in the given context and which preserved the flow of the dialogue as well as the GRE output as much as possible.

3.3 Relevant Attributes

Based on previous work on GRE in the furniture domain (TUNA and COCONUT) we decided to generate referring expressions based on a database that contained the following attributes of the objects used in the study: *type*, *colour*, *size* and *location.* Various studies have shown that people have particular preferences in using these absolute (e.g. *colour*) and relative (e.g. *size*) attributes (e.g. Ford and Olson, 1983, Whitehurst and Sonnenschein, 1978, Pechmann, 1989, Belke and Meyer, 2002). The work on GRE algorithms has generally accepted these findings and applied them to simple domains and artificial contexts.

However, as regards the utility of attributes with respect to purchasing decisions, one could argue that such a simplification will not do. For instance, in the particular situation in which a person wants to buy a chair, there may be attributes, other than *colour* or *size* that are important too. Intrinsic attributes of such a chair are probably the most important, but perhaps not all suitable for identification purposes. A customer is likely to be interested in trying out if the chair is comfortable for her, in measuring its height and width in relation to her own size or the space the chair is intended to occupy at home or work, or in feeling the fabric of the chair to make sure she really likes it etc. In contrast, the actual location of the chair in the shop (which is not necessarily fixed) would be of a secondary importance, because it is not the particular location of the chair within the shop that the customer would be taking home.

For the sake of naturalness, other information about the objects was included in the discourse, but not as part of the REs. For instance, description RE1, in Table 1, was embedded in the dialogue as follows[1]:

> Irish Furniture Seller: ‘A chair which is very comfortable is *the large red one in the front*. It has a nice colour and is not too costly.’

[1] Our dialogue and perception study were set up for native English speakers in Ireland.

In translating English REs to Japanese, two distinct issues are at play: the fact that the utility of the attributes for object identification may not be the same in English and Japanese, and the fact that the translation of the attributes into Japanese might in itself affect the 'human-likeness' of the dialogue, which would obviously be a problem for a cross-cultural study where human-likeness is one of the variables being studied. One way to settle the first issue would be through corpus-based study (cf. Spanger et al., 2009, Theune et al., 2010). The second issue could be approached through text validation studies. Although we acknowledge that these factors might bear on the utility of different attributes across languages and socio-cultural settings as well as affect the naturalness of the translated descriptions, we decided to keep the attributes used in our study as similar as possible to those currently used in GRE research.

3.4 Dimensions of location

Locative expressions are generally used in REs to guide the addressee's eyes to the target object. In the virtual furniture shop each furniture item has a particular (absolute) position and also stands in a particular relation to all other items in the shop. Thus, descriptions RE1, RE2 and RE4 include an absolute locative expression, while descriptions RE3 and RE5 include a relative locative expression. Translation of these English locative expressions into Japanese was problematic. To begin with, there is no straightforward correspondence between English prepositions and Japanese post-positional particles. In English, spatial relations are often represented by prepositions (e.g., 'in', 'above'), whereas in Japanese spatial relations are often represented by spatial nouns and post-positional particles, or by post-positional particles alone (Tokunaga et al., 2005).

In addition, in Japanese, spatial relations are not only dependent on the spatial context, but also on more abstract dimensions such as time and emotion. The Japanese language has its own unique system of referring to things which are near or far to the speaker and the addressee. This system of demonstrative pronouns, adjectives and adverbs, consists of three families of words:

- *the 'a' family of words* is generally used to refer to items that are far away;
- *the 'so' family of words* is generally used to refer to items that are not too near, but not too far;
- *the 'ko' family of words* is generally used to refer to items that are near.

The above is a simplified representation of this system. Not only can the terms ‘near’ or ‘far’ represent distances measured according to space, time and emotion, but the relationship between the addressee and speaker (in terms of whether they both share the same spatial or temporal or emotional perspective) has an influence in the choice of *a* or *so* or *ko*. Hence, the system is dependent on various relative dimensions that may differ per speaker and context. The underlying assumption for the expressions in Table 1 is that the speaker and the addressee, standing side by side in the same time frame, share at least the same spatial and temporal perspectives towards the relevant furniture items. For further discussion of *ko*, *so* and *a*, see (Hasegawa, 2000, Morita, 2002).

With this general knowledge about the *a*, *so* and *ko* system, we note the following about the translations in Table 1: expression RE1 uses a demonstrative pronoun from the *ko* family (‘kochira’) which indicates nearness. Expression RE2 uses a demonstrative pronoun of the *a* family (‘achira’) which indicates a ‘large distance’. Expression RE3 uses a demonstrative adjective of the *so* family which, as explained above, does not refer to a physical distance, but a distance in time (i.e. ‘sono’ refers to the large blue desk that was mentioned earlier in the dialogue). Finally, expression RE4 uses a demonstrative pronoun of the so-family (‘sochira’) indicating a place or direction that is not too close nor too far away from the speaker. Expression RE5 does not use any demonstrative form. Instead it expresses the fact that ‘the red ones’ have been talked about before by the inclusion of a temporal expression ‘ima no’ (‘[you saw] now’) in the utterance right before the RE. Thus the use of a demonstrative was rendered redundant.

Note that although ‘kochira’, ‘achira’ and ‘sochira’ were selected by the translator for the above-mentioned expressions, these demonstrative expressions are substitutable with other expressions. For example, instead of the combination of demonstrative pronoun ‘kochira’ and post-positional particle ‘no’, the demonstrative adjective ‘kono’ or ‘sono’ may be used. Or, for the same example, other demonstrative pronouns that indicate ‘position’, such as ‘koko’ or ‘soko’ can also be used in combination with the post-positional particle ‘no’. In this case, the choice between *ko* and *so* words would depend on the speaker’s judgement of what belongs to the speaker’s own domain, in terms of spatial or temporal or emotional realms. Furthermore, whether to use ‘kochira’ or ‘kono/sono’ or ‘koko/soko’ depends on a ‘politeness’ criteria. Here, for expression RE1, the demonstrative ‘kochira’ was selected based on the assumption that the speaker would feel the target object to be close enough to consider it within his spatial domain, and also with the consideration that the speaker, a shopkeeper speaking in socially polite form, would choose a polite form of demonstrative to convey his message to his addressee, the furniture buyer.

The GRE work on *location* has mainly focussed on the perceptual grouping of objects (Thorisson, 1994, Funakoshi et al., 2004, Kelleher et al., 2005, Funakoshi et al., 2006, Gatt, 2006, Kelleher and Kruijff, 2006), that is on spatial information only. In addition, (Piwek and Cremers, 1996) investigated the use of demonstratives as a comparison of Dutch and English demonstratives in terms of the accessibility of the target and show that English and Dutch speakers follow opposite strategies. Piwek et al. (2008) explain those differences in terms of the use of pointing gestures. To our knowledge, issues of distance and dimensions of time and emotion as they can be indicated with the Japanese *a*, *so* and *ko* have been addressed only by (Byron and Stoia, 2005), who present a motivation for choosing either a proximal or a distal demonstrative based on three dimensions (i.e. spatial, temporal and task performance). Their analysis of a corpus of recorded collaborative dialogues in the English language of participants solving a treasure hunt problem in a virtual space, shows that in English (1) distal demonstrative are used for objects that are located close to or far away from the speaker, whereas proximals are used for objects located near to the speaker; (2) proximal demonstratives are used for objects that relate to the current time and the future, while distals are used for past time; and (3) distal demonstrative are less sensitive to the space and time dimension and more sensitive to the task than proximal demonstratives.

3.5 Representations of location

Translation involves a comprehensive search for the phrase that best matches the meaning expressed in the English original. In determining the appropriate Japanese phrases for absolute locative expressions, subtle semantic similarities and discrepancies between English-Japanese phrases posed problems. For instance, in an attempt to translate 'in the front' as it is used in expression RE1 (i.e. 'in the front of the shop'), it was found that no exact Japanese equivalent exists. The translator had to find an alternative, using the actual situation in the furniture shop. Arguaby, as illustrated in Figure 1, the speaker can see the large red chair as located in front of other objects). Accordingly, in our scenario, the word 'temae', a relative locative expression that depends on the relative position from which the speaker perceives the target object, could be used in the sense of 'located in front of other objects' (cf. Tanaka and Matumoto, 1997). Thus, a combined expression of 'definite article' and 'absolute location marker' in the English language is transformed into a combined expression of a 'speaker's-perception-dependent demonstrative of the *ko* family' and a 'relative location marker' in the Japanese language (i.e. kochira no temae no . . .). There were several approximate translations of 'in the front' as used in description (1). For instance, a noun phrase that includes a demonstrative adjective (e.g., 'sono temae no . . . ') or a demonstrative pronoun (e.g., 'soko no temae no . . . ')[2]. Which combination of words is most preferred by the Japanese speaker seems to be dependent on the context and the natural flow of the dialogue. Translation of the absolute locations in expression RE2, 'in the back' ('achira no oku . . . '), and in expression RE4, 'in the middle' ('sochira no nakahodo . . . '), was handled in a similar fashion.

For the relative locative expressions in expression RE3 and expression RE5, a slight discrepancy of meaning was detected between the seemingly equivalent expressions 'next to' and 'tonari'. The Japanese 'tonari' seems to require a situation in which objects are located so close together that they (almost) touch each other. Hence, initially, in the virtual furniture shop, where there was a visible amount of space between objects, 'tonari' did not seem to apply and a different Japanese expression, meaning 'to the right of', was chosen (cf. Funakoshi et al., 2006). Nevertheless, the actual difference between the two meanings may come from what Europeans comfortably feel as one thing 'next' to another versus what Japanese people comfortably feel as one thing 'tonari' to another. In other words, this difference may arise from the difference of 'sense of physical closeness/distance'. For instance, the furniture shop used for this study may seem spacious to a Japanese person, while it may look cramped to an Irish person.

In the GRE literature the choice between 'next to' and the more specific 'to the right of', has been discussed and implemented in terms of basic level values (cf. Dale and Reiter, 1995, Krahmer and Theune, 2002). Implementation in this particular furniture shop, renders 'next to' for both descriptions RE3 and RE5, because there is no other desk located to the left of the object referred to by the pronoun 'it' (description RE3), and there are no chairs to the left of the 'the red ones' (description RE5). Now, what would producing a basic level value such as 'next to' do for the Japanese addressee? Arguably, it would make the RE more difficult to interpret, because the addressee would have to check both sides of the relatum when interpreting the description. However, one could also argue that producing 'next to' makes interpre-

[2] The particle 'no' usually joins two nouns as in 'A no B' (where A and B are nouns) and causes the meaning of A to modify and restrict the meaning of B.

tation easier, because with ‘to the right of’ it could be unclear which perspective the speaker has chosen: his own, the addressee’s, or maybe the perspective of the relatum. To be able to test the same descriptions across cultures, it was decided to rearrange the furniture so that the setting could appropriately be described by the Japanese expression ‘tonari’.

3.6 Singulars versus Plurals

As illustrated by (Van Deemter and Krahmer, 2006) the graph-based algorithm (Krahmer et al., 2003) can easily accommodate the generation of REs for sets of objects. Hence, we decided to include references to sets of objects (descriptions RE4 and RE5) in our perception study. However, in the Japanese language nouns do not have a plural form. The singular chair of description RE1, and the set of chairs referred to in description RE4, would both be described as ‘isu’ (‘chair’).

Alternatively, a combination of a numeral and a classifier can be used to explicitly state the number of objects in a set (e.g., ‘2 kyaku no isu’ or ‘two chairs’). In the furniture shop using numerals would not be feasible as the shop contains numerous objects that appear to be grouped together. Intuitively, this would make communication seem unnatural, because the furniture seller would first need to count the objects in a group before uttering the RE. Similarly, the furniture buyer would need to count the objects in a group before she could be sure to have identified the correct set.

As the Japanese translation lacks the information that comes from the morphology of the English noun, in our setting the translations of descriptions RE4 and RE5 are ambiguous when considering all objects in the domain as distractors, even when all known attributes are included (i.e., ‘large red chair(s) in the middle’ and ‘small green chair(s) next to the red one(s)’ respectively). Hence, although the GRE algorithm would terminate successfully when generating the English REs, it would result in failure when generating the Japanese REs due to the lack of cardinal information.

A possible way to add a sense of plurality to the Japanese translations of descriptions RE4 and RE5 would be to enhance the locative expression as in ‘the [large red / small green] chairs grouped in the middle’. Adding ‘grouped’ to the locative expression would indicate that there are a number of chairs. However, this would require the knowledge base used by the GRE algorithm to include some information on perceptual groupings (Kelleher et al., 2005, Kelleher and Kruijff, 2006).

As regards our cross-cultural perception study, one might question if the same algorithmic output is tested when a part of the distinguishing attribute values of the REs cannot be equivalently realised in the languages under consideration. For GRE evaluation purposes, simply using similarity metrics which focus on semantics rather than realisations, as is the current practice, seems insufficient to capture the differences between bi-lingual pairs of REs.

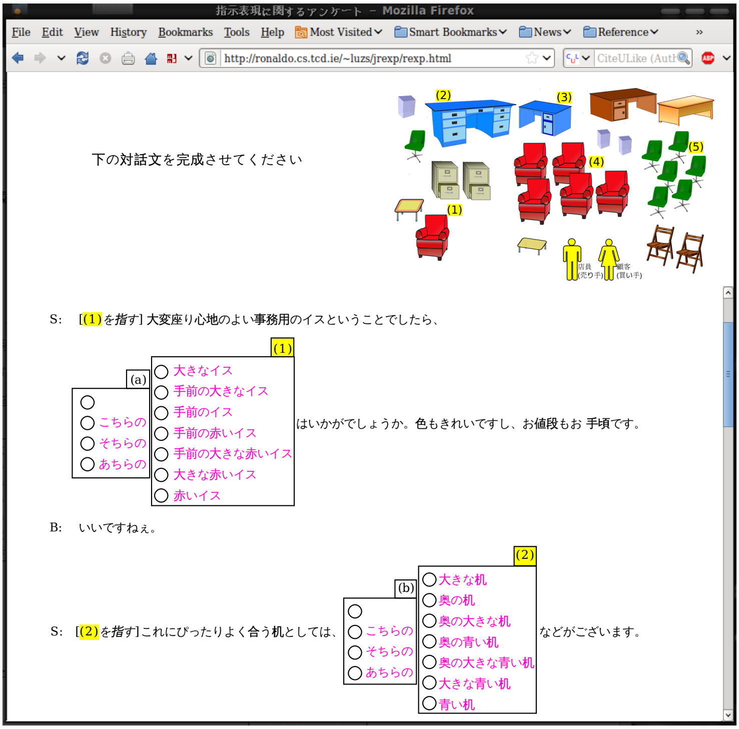
4 Two Elicitation Studies

The observations made in the translation process considered the full potential of the five REs in our dialogue, i.e., including demonstratives and all available properties. In this section, two data elicitation studies are presented which were conducted to find out what type of REs native speakers of Japanese would prefer to use in the dialogue, given the set of available properties (*colour*, *size* and *location*) and given the multimodal setting in which the actors would point from their stationary position in the direction of the target objects. The studies used the same materials but different methods: Study I was a qualitative focus group study conducted with six native speakers of Japanese in a lab-based setting, while Study II was conducted over the internet and thus rendered much more quantitative data.

4.1 Materials

The study was presented to the participants through a web browser and consisted of three pages. The first was a tutorial page in which the participants could were informed about the goals of the study, what they were going to see on the next page, and what we asked them to do. The second page is illustrated in Figure 2. At the top of the screen a picture of the domain was presented. The bottom part of the screen contained the dialogue through which the participants could scroll and select the REs they preferred from a set of options, all of which were simultaneously available to the participant while reading the sentence. The picture of the domain was always visible on the top part of the screen while participants scrolled through the dialogue.

The five relevant REs were each presented with two boxes as illustrated in Figure 3. One, the DE-box, in which participants could select a determiner or demonstrative, and the other, a RE-box, in which a RE could be chosen. As the Japanese language does not have any determiners the DE-box had an empty option which allowed the participant to leave that position blank. The other three options corresponded to the *a*, *so* and *ko* forms. The RE-box contained seven possible REs in which the inclusion of *colour*, *size* and *location* were varied; all REs contained the relevant value for *type* as a noun. For instance, in the case of the second RE in the dialogue, the options would be the Japanese equiv-


指示表現に関するアンケート – Mozilla Firefox
File Edit View History Bookmarks Tools Help Most Visited Smart Bookmarks News Reference »
http://ronaldo.cs.tcd.ie/~luzs/jrexp/rexp.html
CiteULike (Auth
(2)
(3)
(4)
(5)
(1)
店員
(売り手)
顧客
(買い手)

下の対話文を完成させてください

S: [(1)を*指す*] 大変座り心地のよい事務用のイスということでしたら、

(a)
○
○ こちらの
○ そちらの
○ あちらの

(1)
○ 大きなイス
○ 手前の大きなイス
○ 手前のイス
○ 手前の赤いイス
○ 手前の大きな赤いイス
○ 大きな赤いイス
○ 赤いイス

はいかがでしょうか。色もきれいですし、お値段もお 手頃です。

B: いいですねぇ。

S: [(2)を*指す*]これにぴったりよく合う机としては、

(b)
○
○ こちらの
○ そちらの
○ あちらの

(2)
○ 大きな机
○ 奥の机
○ 奥の大きな机
○ 奥の青い机
○ 奥の大きな青い机
○ 大きな青い机
○ 青い机

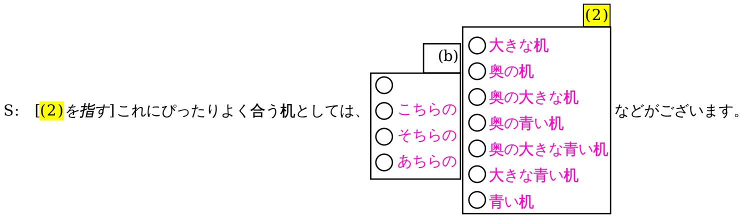
などがございます。

FIGURE 2 Sreenshot of the second page of the web experiment on which participants were asked to choose their preferred REs, where two *S*eller and one *B*uyer utterances, DE-boxes (*a* and *b* and RE-boxes *1* and *2* are visible.

alents of 'large desk', 'blue desk', 'desk in the back', 'large blue desk', 'large desk in the back', 'blue desk in the back' and 'large blue desk in the back'. After each RE-box, it was stated that the agent's utterance of the RE would be combined with a pointing gesture in the direction of the target. The third webpage consisted of a thank you note and information on a prize draw as a reward for participating in the study.

4.2 Hypotheses

The hypotheses for the REs to be selected by the participants in our two studies are based on findings from cognitive linguistics (Pechmann, 1989, Arts et al., 2010, To appear) which show that absolute properties (e.g. colour) are preferred to relative properties (e.g. *size*). Following Krahmer and Theune (2002) we expect that locative expressions are even less preferred than relative properties. Recall that the Van der



S: [(2)を指す]これにぴったりよく合う机としては、

(b)

○

○ こちらの

○ そちらの

○ あちらの

(2)

○ 大きな机

○ 奥の机

○ 奥の大きな机

○ 奥の青い机

○ 奥の大きな青い机

○ 大きな青い机

○ 青い机

などがございます。

FIGURE 3 An example of how the REs were presented, showing the *S*eller's utterance with a first-mention RE to referent 2 including the DE-box *b* and the RE-box *2*.

Sluis & Krahmer algorithm determines the content of the RE based on a notion of effort formalised as a cost function, which ensures that the algorithm searches for the least costly solution. The human preference order of the properties can be instantiated through a definition of the algorithm's cost function so that *location* is the most costly attribute, *size* is cheaper, and *colour* is the least costly.

In our set up, we presented the discourse domain including the agents that featured in the dialogue in a two-dimensional fashion. However, we asked the participants to imagine that the furniture seller agent included a pointing gesture to accompany the linguistic descriptions to refer to the target objects. Hence we asked participants to consider the distinguishing effect that this pointing gesture would have in a three-dimensional environment. As we cannot be sure about the scope of these pointing gestures in the minds of the participants and their effect on the distractor set of objects on which the participants based their choice of RE, we decided to test two hypotheses for the type of REs collected.

Our first hypothesis, H1, considers all objects in the domain as depicted in Figure 2 apart from the target as distractor objects (i.e., the pointing gesture has no effect, it does not rule out any distractors). Accordingly, for RE1 the algorithm will first include *colour* to rule out all objects in the domain that are not red. For RE1, *size* will not remove any distractors and is therefore not added to the RE. The property *location*, however is included to rule out the group of red chairs in the middle of the shop. For RE2, the algorithm selects *colour* to rule out all objects that are not blue. Secondly, *size* is added to RE2 to remove the remaining smaller blue desk and thereby empty the set of distractors. For RE3, the algorithm adds *colour* to rule all distractors that are not blue and adds the property *size* to remove the large blue desk from the distractor set and uniquely distinguish RE3. RE4 is distinguished by first adding *colour* and ruling out all objects that are not red. Then *location* is added to remove the only remaining distractor, that is the singular red chair in the front of the shop. RE5 is built by adding *colour*, which leaves a distractor set with only green objects. Subsequently *location* is added to remove the singular green chair on the left-hand side of the domain.

Our second hypothesis, H2, considers only the set of distractor objects located in the scope of the pointing gesture performed by the agent to distinguish the target object. For all five targets we tentatively defined the scope of the pointing gestures as depicted in Figure 4, where the areas covered by the pointing gestures are of the same size, but differ in terms of the target which is located in the center of the gesture's scope.

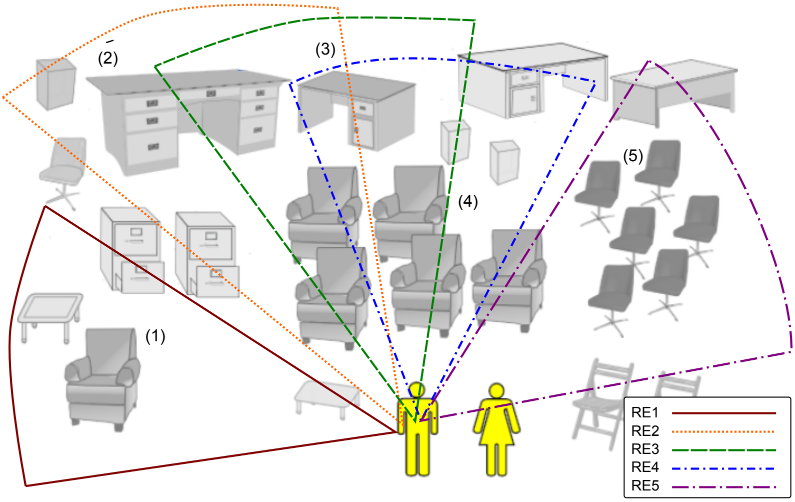


FIGURE 4 Furniture shop divided into five areas that cover the scope of the pointing gestures produced by the Seller agent to accompany the REs *RE1* to *RE5*.

For the sake of illustration, we define the set of distractor objects as including all objects (i.e. excluding the target object) that are located fully or partly within the projected lines that indicate the scope of the gesture. In addition, pointing gestures will be considered to be less costly then any of the linguistic attributes.[3] Therefore, for all five REs the algorithm will add a pointing gesture first which results in

[3] As mentioned in Section 2 we will not consider the relative cost of a pointing gesture in this paper, but just add it to the RE

a decrease of the distractor set. For all five REs, however, inclusion of the pointing gestures does not result in distinguishing REs and the algorithm still needs to add linguistic information to identify the targets uniquely. For RE1, the algorithm adds *colour*, with which the distractor set can be emptied (i.e. the pointing gesture had already ruled out the group of red chairs in the middle of the shop). For RE2, the algorithm adds *colour* and *size*; the pointing gesture's scope has decreased the target set but still includes some objects with a different colour as well as the smaller blue desk. RE3 is also extended with *colour* and *size* to respectively rule out the objects in the gesture's scope that are not blue as well as the large blue desk. Both RE4 and RE5 are enriched with *colour* which removes the remaining distractors that are located in the scope of the respective pointing gestures.

With a lack of further reference on Japanese determiners and demonstratives, we expected the preferences to be the realisations selected by our translator (H3). It should also be remarked that we expect the participant's choice of determiners to affect their choice of RE attributes. In particular, we expect this to be the case for the *location* attributes, since the determiners are known to convey proximity information. Another factor that should affect both the choice of determiners and, perhaps indirectly, the choice of attributes is the order in which REs are generated in the dialogue. It is expected, for instance, that by referring to object (2) with a determiner that implies an object located far from the speaker, the speaker brings object (3), which is located next to object (2), somewhat "closer" to the speaker and hearer, in the more abstract sense explained in Section 3.4.

In summary, we test the following three hypotheses for the generation of Japanese REs that include a pointing gesture:

H1: Participants consider all objects in the domain as distractors from which the target has to be distinguished.

H2: Participants consider only objects in the scope of the pointing gesture as distractors from which the target has to be distinguished.

H3: Participants agree with our translator with respect to the choice of determiner.

These hypotheses are further specified in Table 2, in terms of the attributes and determiners the participants are expected to choose.

In order to test hypotheses H1 and H2 we compared the realised output of our GRE algorithm with the participants' RE preferences. We chose the Dice coefficient as our evaluation metric, which accounts for a degree of overlap between two descriptions. Dice computes the degree of similarity between two sets by scaling the number of attributes that

TABLE 2 Expected REs for referents *RE1* to *RE5* for three hypotheses *H1* and *H2* on the content of the REs and *H3* on the choice of determiner.

Target	H1: Whole Domain	H2: Gesture Scope	H3: DET
RE1	colour, location	colour	'kochira'
RE2	colour, size	colour, size	'achira'
RE3	colour, size	colour, size	omitted
RE4	colour, location	colour	'sochira'
RE5	colour, location	colour	omitted

the two descriptions have in common according to the overall size of the two sets, as shown in equation (1.1):

$$dice(H_a, R) = \frac{2 \times |H_a \cap R|}{|H_a| + |R|} \tag{1.1}$$

where H_a is the set of attributes in the description produced by a human author, and R the set of attributes in the reference description generated by the algorithm.

Dice yields a value between 0 (no agreement) and 1 (perfect agreement). The attributes are chosen from a set $A = \{c, s, l\}$, denoting *colour*, *size* and *location*, respectively, so that possible H_a will be elements of $\mathcal{A} = 2^A \setminus \emptyset$. We summarise the Dice scores by their expected values for a particular object. That is, we report the mean scores weighted according to the probability p_a that a combination of attributes $a \in \mathcal{A}$ is chosen, as set out in equation (1.2).

$$E[dice(H, R)] = \sum_{a \in \mathcal{A}} p_a \times dice(H_a, R) \tag{1.2}$$

As an indicator of the overall recall of our algorithm, we will also report the 'perfect recall percentage',(PRP), that is the proportion of times the algorithmic output matches the participant's choices exactly. For H3 we will only discuss the PRP scores.

5 Study I: Elicitation by Focus Group

Participants

Six researchers at the National Institute of Informatics at Tokyo participated in Study I, all were native speakers of Japanese. One of the participants was female and five were male. Four of the participants were students between 20 and 30 years old, and two were academics, one between 31 and 40 and the other between 41 and 50 years old.

Procedure

The participants and the experimenter met in a lab, where the pages of the web experiment were projected on a wall-mounted 65 inch screen. The participants took seats that allowed them a good view on the screen. The experimenter gave a short introduction of the study after which she referred the participants to the first page of the web experiment as discussed above. When the participants indicated that they had finished reading, the experimenter moved on to the second page of the experiment in which the participants were asked to select demonstratives and REs. For each reference, each of participants chose an option from the DE-box and one from the RE-box in silence and wrote them down hidden from the other participants. Per reference, when all participants had made their choice, the experimenter collected the results in a plenary fashion, and subsequently the participants discussed their choices. As the experimenter did not speak Japanese and not all attendees were able to speak English, the attendees discussed the material amongst themselves in Japanese and then translated back to the experimenter. Throughout the session, the experimenter scrolled back and forth through the dialogue when asked to do so by the group.

Results

Tables 3 and 4 present the choices of REs, determiners and demonstratives by the participants of Study I.

TABLE 3 Percentages (and frequencies between brackets) of REs collected with Study I for *RE1* to *RE5* for which the values of the available attributes *colour*, *size* and *location* are indicated, as well as the actual choices made by the participants in the study as combinations of *c*olour, *s*ize and *l*ocation.

	RE1	RE2	RE3	RE4	RE5
colour, *size,* *location*	*red,* *large,* *front*	*blue,* *large,* *back*	*blue,* *small,* *next*	*red,* *large,* *middle*	*green* *small,* *next*
c	50% (3)	–	–	33% (2)	67% (4)
s	–	–	–	–	–
l	–	17% (1)	–	–	–
cs	17% (1)	–	–	17% (1)	17% (1)
cl	17% (1)	50% (3)	–	33% (2)	17% (1)
sl	–	–	17% (1)	–	–
csl	17% (1)	33% (2)	83% (5)	17% (1)	–

TABLE 4 Percentages (frequencies shown in brackets) of the number of times a demonstrative of the *a*, *so* and *ko* family, or 'no determiner' was selected by participants in Study I for each of the five referring expressions, RE1 to RE5.

	'a'	*'so'*	*'ko'*	no det
RE1	100% (6)	–	–	–
RE2	100% (6)	–	–	–
RE3	33%(2)	50%(3)	–	17%(1)
RE4	–	17%(1)	83%(5)	–
RE5	33%(2)	50%(3)	17%(1)	–

During the lab session, the participants discussed their choices. About RE1 (i.e., 'the large red chair in the front') it was noted that the locative expression 'in the front' was troublesome. Four out of six participants found the Japanese realisation of that property as 'temae' inappropriate because from the seller's point of view there were no objects located behind the chair. Yet, two of the participants had chosen to include 'temae' in the description. Also for RE2 (i.e., 'the large blue desk in the back') the locative expression 'oku' (i.e. 'back') which was included by 83% (5) of the participants was considered somewhat ambiguous. For RE3 (i.e., 'the small blue desk next to it') the participants were divided about the demonstrative. The majority found that, in contrast to a demonstrative of the *a* family (i.e. 'far away'), a demonstrative of the *so* family (i.e., 'not near and not far') would establish a connection with the reference to RE2 earlier in the dialogue. Participants found the locative expression used in RE4 for 'middle' ('nakahodo') too "classical", instead they proposed to use the form 'center' ('Ma n naka'), or 'temae'. Still, it was agreed that a locative expression and/or a gesture was definitely needed to distinguish the target set of RE4 from the target of RE1. For RE5 the relative locative expression 'tonari' in 'next to the red ones' was found too vague, again the majority of the participants would have preferred the use of 'temae' ('in the front').

Table 5 presents the Dice scores and PRP scores for the data collected in Study I. The Dice scores for the collected REs in Study I are all above .5. The scores for H1 (i.e. with the objects in the whole domain as distractors) and the scores for H2 (i.e., taking only the objects in the scope of the pointing gesture as distractors) are very similar. However, the PRPs (except for RE2 and RE3, which are identical under H1 and H2) are higher for hypothesis H2, which indicates that a better match with human produced REs can be obtained when only the objects in

TABLE 5 Dice scores and Perfect Recall Percentages *PRP* for the REs collected in Study I

	H1-Dice	*H1-PRP*	*H2-Dice*	*H2-PRP*	*H3-PRP*
RE1	.72	17%	.70	50%	0%
RE2	.56	0%	.56	0%	100%
RE3	.76	0%	.76	0%	17%
RE4	.78	33%	.71	33%	17%
RE5	.70	17%	.86	67%	0%

the scope of the pointing gesture are considered as distractors.

Discussion

Although still ambiguous due to the fact that the Japanese language does not distinguish between singular and plural descriptions, our participants' preferences for RE1 and RE5 seem more in line with H2 than with H1, that is respectively 50% and 67% agreed with H2 (i.e. taking into account the reducing effect of the pointing gesture on the distractor set), versus respectively 17% and 17% that agreed with H1 (i.e. considering all objects in the domain as distractors). For RE4 preferences were equally divided: 33% agreed with H1 and 33% agreed with H2. In the chosen descriptions for RE1, RE4 and RE5 the targets of RE1 and RE4 could be confused with each other as they are all red chairs and the target set of RE5 could be confused with the singular green chair located at the left-hand side of the domain. For RE2 and RE3 our hypotheses H1 and H2 are the same, including both *colour* and *size*. There are no exact matches. The majority of descriptions chosen for RE2 (67% in total between 'l' and 'cl') are ambiguous and even when taking into account the scope of the pointing gesture it would be necessary to include the target's size to disambiguate them. In contrast, for RE3 (i.e. 'the small blue desk next to it') a more redundant description is preferred by 83% of the participants, where including a locative expression where *size* and *colour* would have sufficed. For the determiners we found 100% agreement with our translator for RE2, where 'achira' indicates that the object is far away from the speaker. For the other REs the determiner choices greatly differed from what our translator had proposed.

Although this small focus group study by itself cannot decide the status of the hypotheses, it is worth mentioning that participants displayed a great variety in their preferences of the realisations for almost all REs (except perhaps RE3). In addition, the participants did not seem to make any special effort to avoid ambiguity in the REs they chose. In one case (RE2), ambiguity was, in fact, highly preferred.

In general, participants seemed to find it difficult to choose from the available options. It often happened that they suggested a different phrase. Participants expressed that they sometimes felt that the form of REs was dependent on what had happened earlier in the discourse. It may have been a handicap that the participants were not able to scroll through the dialogue themselves but had to ask the experimenter to do this for them.

In the discussions during and at the end of Study I, we found two factors that may have caused the large variety in choice of determiners and demonstratives. Firstly, participants remarked that the two-dimensional sketch of the domain made it difficult to define the space of the shop and the distances between the agents and the furniture items. A second factor could be that the participants only realised during the final discussion in Study I that the DE-box could be left blank. Perhaps this was not stated clearly enough in the introductory webpage of the experiment, but it also could have been an effect of the experimenter controlling the experiment pages such that the participants were not able to try out the available choices themselves.

6 Study II: Elicitation through a Web Experiment

Participants

The data used for this part of the study were collected from a total of 62 participants. Of those, 56 were native speakers of Japanese, and their answers comprise the data analysed below. About 26% (16) of them were female, 73% (45) male and 1% (1) left their gender unspecified in the form. unspecified. As regards occupation, 50% (31) of the participants were students, 13% (12) were academics and 37% (23) categorised their occupation as ‘other’. Just under 55% (34) of the participants were between 20 and 30 years old, 27% (17) were between 31 and 40 years old, 16% (10) between 41 and 50, and 1 participant did not specify his/her age. Note that the Japanese version of the dialogue script was tailored to a particular specification of the actors and their context was independent of the audience of the script.

Procedure

The experiment website was distributed through sending invitations for participation by email to six acquaintances in Japan with a request to pass the invitation on to other native speakers of Japanese. Participants took part in the study at their own time and in their own pace. After reading through the tutorial page of the experiment, they could click a button at the bottom of the page to proceed with the study. Upon completing their choices participants were offered the opportunity to enter free-form comments in a text box and asked to state their gender, age group, occupation and whether they were native speakers of Japanese. Once the form was submitted, the partipants were thanked for their efforts and informed that they had been included in the prize draw.

Results

Table 6 shows that, in the references to the chairs (RE1, RE4 and RE5), a majority of the participants (between 40% and 50%) preferred a RE consisting solely of a *colour* attribute. The second most preferred choices for RE1 and RE5 were the combined use of *colour* and *size* (28.6% in both cases), while for RE4 it was *location* and *colour*. References to the desks were generally more evenly divided. For RE2, participants mostly decided to include all properties available (37.5%), with the combination of *colour* and *size* as the second most popular choice. For RE3 including both *size* and *location* was preferred (30.4%)[4], followed very closely by including all properties (28.6%). The Dice scores in Table 7 show that, for expressions with different hypotheses (RE1, RE4 and RE5), H2 received higher scores than H1. The differences for RE1 and RE5 were found to be significant at the $p < .05$ level (t-test t[95]=2.1 and t[103]=3.2, respectively).

TABLE 6 Frequencies of REs collected with Study II for *RE1* to *RE5* for which the values of the available attributes *colour*, *size* and *location* are indicated, as well as the actual choices made by the participants in the study as combinations of *c*olour, *s*ize and *l*ocation.

	RE1	RE2	RE3	RE4	RE5
colour, *size,* *location*	*red,* *large,* *front*	*blue,* *large,* *back*	*blue,* *small,* *next*	*red,* *large,* *middle*	*green* *small,* *next*
c	42.9% (24)	7.1% (4)	5.4% (3)	46.4% (26)	48.2% (27)
s	8.9% (5)	14.3% (8)	8.9% (5)	3.6% (2)	8.9% (5)
l	1.8% (1)	5.4% (3)	5.4% (3)	1.8% (1)	0.0% (0)
cs	28.6% (16)	19.6% (11)	7.1% (4)	12.5% (7)	28.6% (16)
cl	5.4% (3)	5.4% (3)	14.3% (8)	23.2% (13)	7.1% (4)
sl	3.6% (2)	10.7% (6)	30.4% (17)	3.6% (2)	0.0% (0)
csl	8.9% (5)	37.5% (21)	28.6% (16)	8.9% (5)	7.1% (4)

[4] Note that this is not a possible output of our algorithm.

TABLE 7 Dice scores and Perfect Recall Percentages *PRP* for the REs collected in Study II.

	H1-PRP	*H2-PRP*	*H3-PRP*	*H1-Dice*	*H2-Dice*	*H1 vs H2*
RE1	5.4%	42.9%	41.1%	.58	.70	*
RE2	19.6%	19.6%	69.6%	.72	.72	
RE3	7.1%	7.1%	10.7%	.62	.62	
RE4	22.2%	46.4%	16.1%	.71	.75	
RE5	7.1%	48.2%	3.6%	.59	.76	*

* denotes significant difference at the $p < .05$ level

With the exception of the preferred references to the object furthest away from the agent (RE2 and RE3), for which H1 and H2 are the same, PRPs are also higher for H2, indicating that participants most often took into account the scope of the agent's pointing gesture directed to the target.

In most cases, excepting RE1, there was a clear majority ($\chi^2[12] = 127.96, p < .05$) for a particular choice of demonstrative, as shown in Table 8. For RE1 participants were divided between a demonstrative of the *a* family (i.e. 'far away'; 39.3%) and *ko* family (i.e. 'near by'; 41.1%). The target of RE2 (i.e. 'the large blue desk in the back') is mostly considered to be far away (69.6%). However, 50% prefer a demonstrative from the *so* family (i.e., 'not near and not far') for the smaller desk next to it (RE3), with *a* coming second (33.9%). The set of red chairs in the middle (RE4) is mostly considered near to the speaker (69.9%), while 50% of the participants chose a *so* demonstrative for the green chairs (RE5). The Dice and PRP scores presented in Table 7 show some agreement with the demonstratives suggested by our translator for RE1 (41.1%) and RE2 (69%).

TABLE 8 Frequencies of the number of times a demonstrative of the *a*, *so* and *ko* family, or 'no determiner' was selected by participants in Study II for each of the five referring expressions, RE1 to RE5.

	'a'	*'so'*	*'ko'*	no determ.
RE1	39.3% (22)	17.9% (10)	41.1% (23)	1.8% (1)
RE2	69.6% (39)	17.9% (10)	1.8% (1)	10.7% (6)
RE3	33.9% (19)	50.0% (28)	5.4% (3)	10.7% (6)
RE4	8.9% (5)	16.1% (9)	69.6% (39)	5.4% (3)
RE5	32.1% (18)	50.0% (28)	14.3% (8)	3.6% (2)

As regards possible interactions between choice of determiner and

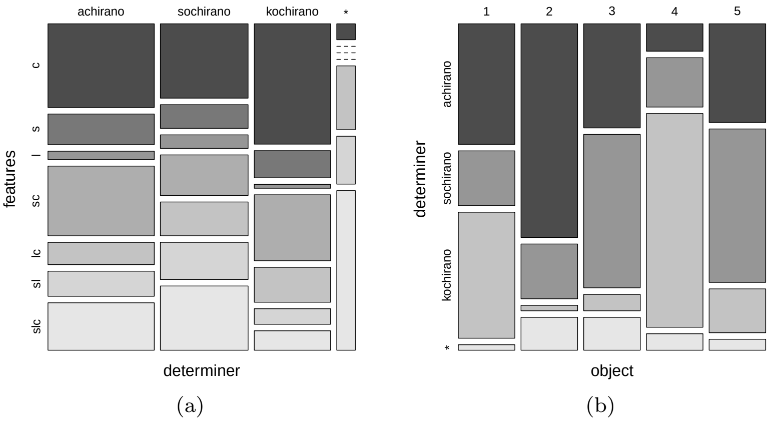


FIGURE 5 Mosaic plots showing (a) the distribution of determiners per feature set and (b) the distribution of determiners per object, in Study II. '*' stands for 'no determiner'.

choice of attributes, and order of appearance of RE in the dialogue and choice of determiner, our expectations were also partly confirmed. Figure 5 shows mosaic plots representing the contingency tables for determiner by feature set and object. Comparing the determiners to features (Figure 5(a)) we see that the preference for the *colour* attribute mentioned above is stronger for the *ko* family, and that the choice of *ko* seems to imply less need for the *location* attribute. For the *a* and *so* families, attribute sets were employed more often. It is also worth noting that when no determiner is used, participants preferred to use the full set of attributes in describing the object. The choice of determiner was also observed to significantly affect the decision to include an individual attribute in the RE ($\chi^2[6] = 13.41, p < .05$). These observations are confirmed by the estimated probabilities of using a *location* attribute given a particular choice of determiner: $p(l|a) = .36$, $p(l|so) = .51$, $p(l|ko) = .25$ and $p(l|*) = .94$. It may seem somewhat surprising that *so* attracted more *location* attributes than *a* (.51 versus .36). However, this is explained by the fact that, the reference to (3) was made immediately after the reference to (2), which caused the participants to use *ko* for the former, as shown in Figure 5(b), even though the two objects are located roughly at the same distance from the speaker.

The experiment successfully engaged the participants, with half of them leaving comments in the text box that was provided. There were a number of comments related to the sets of linguistic descriptions that the subjects were asked to choose from: one participant indicated that he had preferred to use *size* instead of *colour* to avoid misunderstanding in case the addressee were colour blind. Another participant found that the dialogue history helped to avoid repetitive use of similar descriptions by allowing for omission of particular information. Two participants commented that it was not always easy to make a forced choice and they suggested a different experiment format which would enable participants to write in their preferred REs.

Another set of comments addressed the presentation of the setting in the furniture shop in terms of space and distance: five participants commented on the fact that it was difficult to get a sense of the distances between the agents and the objects in the shop; one of them suggested that the icons that indicated the agents could have been bigger to illustrate that they were located in the front of the shop. One participant thought that it would be unnatural to have a seller and a buyer standing stationary in front of the items for sale. One participant observed that it would not be easy to single out an object with a pointing gesture. One participant found that the timing of the linguistic description and the pointing gesture was left vague.

Summary

In general, the web-based study resulted in data that showed a clear preference for one of the presented REs in each of the five cases, though opinions were more divided when describing objects located further away from the agent. The data present a fairly clear cut decision on how to define the distractor set for generating linguistic descriptions when using the Dice metric for comparing algorithmic and human output. Looking at the percentage of perfect matches for the three REs where the hypotheses differed (RE1, RE4 and RE5), it turns out that H2 results in significantly better scores for RE1 and RE5, indicating that Japanese speakers do consider a reduction of the distractor set as a result of a pointing gesture. As regards H3, participants in Study II also show some agreement in their preferences for demonstratives, but seem to differ considerably from the choices our translator suggested.

7 Conclusions and Future Work

This paper presented an analysis of a number of linguistic issues that are at play when one wants to generate REs in the Japanese language, in particular if attempting to preserve similarities with the properties of REs generated for English. The approach taken to address Japanese REs was to translate a dialogue from English to the Japanese language, where the dialogue contained REs that included values for attributes commonly used in GRE research. Subsequently, two data elicitation studies were conducted to find out if native speakers of Japanese agreed with two particular outputs of the Van der Sluis & Krahmer algorithm as realised in terms of the translation. The two outputs differed in the definition of the distractor set used by the algorithm, either considering all objects in the domain as distractors or only the objects located in the scope of a pointing gesture that would be directed to the target referent.

7.1 Lessons from the translation process

The first-mention REs considered in the English to Japanese translation were part of a dialogue script situated in a virtual, but life-like setting in which two agents, a seller and a buyer, are discussing furniture. The goal of the translation was to adapt the scenario to the Japanese style, while keeping the REs as similar as possible to the English originals. The REs were only five in number but covered the main areas currently researched in GRE, namely, cardinality, locative expressions and the position of the referent in the domain of conversation. A detailed analysis of the challenges we came across in the translation process reveals a number of issues relevant to the work in GRE, especially the work that targets life-like contexts:

- Not all ingredients of English REs have an exact equivalent in the Japanese language. In the translation process, it was obvious from the start that the meanings of English definite articles (e.g. 'the'), anaphoric expressions like 'one' or 'ones' and pronouns like 'it' and plural noun forms (e.g., 'chairs') had to be captured by other means or omitted completely. While these issues, apart from cardinality, do not bear on the choice of descriptive attributes (arguably the core of a GRE algorithm) they do have implications to how the expressions are ultimately realised.
- The discourse context plays an important role, also for first-mention REs. In contrast to English, the Japanese language has a rich vocabulary to express distance in terms of, not only space, but also time and emotion. These dimensions seem to be of particular importance when using REs in a specific context or discourse.
- Specification of locative expressions is not straightforward. The distinction between 'absolute' and 'relative' locative expressions as observed in English (e.g. 'in the front' versus 'next to it' respectively) does not have an exact equivalent in Japanese, where any locative expression is strongly related to discourse factors (e.g., position of the speaker, position of addressee, position of distractor objects).

- The choice of attribute values greatly depends on the visual context. For locative expressions, basic level values like 'tonari' (i.e. 'next to') apply to very specific situations in which two objects are almost touching each other. When this is not the case, a more specific value (e.g. 'to the right of') is chosen. These differences may arise from cross-cultural differences between the 'sense of physical closeness/distance'.
- The relevance or salience of the attributes used in a RE may be dependent on a (scenario-specific) utility function, which is likely to go beyond the usual *colour*, *size* and *location* representations that are commonly used in GRE. Moreover, this function may not be the same across languages.

7.2 Eliciting Japanese REs

In our two elicitation studies, in which native speakers of Japanese were asked to select REs composed from the expressions produced by our translator, we found that participants were not always happy with the realisations they were offered. However, we found a considerable agreement in their preferences as well as a relatively high percentage of exact matches with the algorithm's output that used a reduced distractor set due to the influence of the agent's pointing gesture. With respect to the demonstratives participants were asked to choose for each of the REs, we did not find much agreement with the suggestions of our translator. This may have been due to diferences between the way the scene and setting were described to the translator (verbally, in great detail) and presented to the participants (pictorially, with most details left unspecified).

7.3 Consequences for GRE Research & Future work

The above findings raise the issue of how transferable between languages and cultures existing GRE algorithms are. It seems that existing approaches to GRE make particular assumptions about the target language, such as information carried by a determiner. From an engineering perspective, a question arises as to whether one should model more generic algorithms, possibly by introducing more detailed knowledge representations, or whether it is more beneficial to simply rely on translators to produce REs. Based on our observations we suggest the following directions for further research in multilingual GRE:

- As pronouns (e.g. 'it') and definite articles (e.g., 'one') do not have a counterpart in the Japanese language, approaches that simply check for previous occurrences of the head noun to determine whether to generate a pronoun, as suggested by Dale (1992) and Krahmer and Theune (2002) (see Section 3.2) do not seem to apply to the generation of Japanese REs. Although we cannot make any strong claims based only on the three instances in which we employed these applications, our results indicate that the issues of adapting rule-based approaches to languages such as Japanese deserves further research.

- Japanese nouns do not include any morphological information on cardinality. That is 'chairs' and 'chair' would both be translated as 'isu'[5]. Consequently, GRE algorithms like the ones proposed by (Van Deemter, 2002, Van Deemter and Krahmer, 2006) could render a distinguishing English description, while for Japanese the description would still be ambiguous. A solution to this problem may be to extend or refine the domain representation the GRE algorithm uses with a representation of sets of objects (including singleton sets) perhaps on the basis of perceptual groupings (cf. Thorisson, 1994).
- The Japanese system of demonstrative pronouns, adjectives and adverbs for referring to objects which are near or far to the speaker and the addressee was found to be quite complex in the translation process. Morover, our data elicitation studies did not show a large consensus in how the *a*, *so* and *ko* families are used within the short dialogue we asked the participants to consider. For English, on the other hand, the use of determiners and demonstratives is often thought of as straightforward, i.e., 'this' and 'these' for things that are near and 'that' and 'those' for things that are faraway (cf. Bühler, 1934, Clark, 1996). However, when including the use of precise and imprecise pointing gestures (cf. close and distant pointing gestures as presented by Clark and Bangerter, 2004, Bangerter, 2004)), things become more complicated. The faraway/nearby distinction is predicted by the data gathered by Piwek et al. (2008), but suggests that proximals are mainly used indexically and distals can also have a non-indexical use (i.e. more similar to the use of definite determiners). Overall, GRE research has not addressed the choice of determiners and demonstratives satisfactorily and it would be interesting to further investigate the effects of physical and emotional distance as well as distance in time, in monolingual as well as multilingual settings perhaps taking the approach taken by Byron and Stoia (2005) as a starting point.
- Especially in situated dialogues where speakers are a physical part of the discourse domain, choice as well as realisation of absolute or

[5] The use of a demonstrative form of the *a*, *ko* or *so* family would also not resolve ambiguity in such cases as it could in English (e.g., 'this sheep', vs. 'these sheep').

relative locative expressions seem to require a more specific definition of the domain in terms of the position and roles of the entities included (e.g., position of the speaker, addressee target and distractors). In addition, some possibly cultural dependent or even user dependent (cf., the variations in preferences for demonstratives in Study II) calculations of what should be considered near or far may be called for.

- Although more elaborate approaches exist (Tenbrink, 2004), in GRE objects are often determined to be spatially related when they are located within a "small distance" of each other (e.g. Horacek, 1995, Krahmer and Theune, 2002), where "small distance" is usually interpreted in such a way that two objects are spatially related only if there is no other object located in between these two objects. However, for the Japanese word 'tonari' this would not be specific enough. Similarly, the choice of attribute values in terms of basic-level and more specific values for attributes as proposed by (Dale and Reiter, 1995), may be very much dependent on the visual context. For instance, when several objects in the domain are basically 'red', the different shades of red in the domain may become more important. Perhaps the simple attribute-value representation of properties in the knowledge bases as currently used by GRE algorithms does not suffice when we want to apply such algorithms to more realistic visual contexts. Again, visualisation techniques could be a key to interpret the domain of conversation in order to inform knowledge representation.
- Generally, GRE algorithms employ results from studies of attribute saliency when available. These studies, however usually address simple domains in terms of the number and type of target objects and address often only choice and perhaps syntactic ordering of attributes. In addition, these studies tend to be limited to the English and Dutch languages.

In addition to the more specific investigations suggested thus far, an important future effort needed for GRE research is to collect corpora in languages other than English (cf. Koolen et al., 2009, Spanger et al., 2009, De Lucena et al., 2010). We are currently working on a multilingual replication of the elicitation experiment for English, Dutch and Portuguese in order to find out if the findings presented in this paper can be attributed solely to the characteristics of the Japanese language or may have also been influenced by the context of the experimental set up. We would also like to further explore the significant gender differences obtained in the Japanese data, which we were not able to report here due to space constraints. In addition, we acknowledge that preference experiments like the ones presented in this paper cannot provide the ultimate answer for GRE, especially not for dialogue contexts which, for instance, often include repairs or clarifications to resolve ambiguities between speakers and hearers (Clark and Wilkes-Gibbs, 1986, Carletta et al., 1997). Therefore, we argue for a more holistic approach, which includes extrinsic and qualitative analysis in addition to preference experiments to assess the quality of GRE algorithms in multilingual, multimodal and interactive contexts.

Acknowledgments

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localisation (www.cngl.ie) at Trinity College Dublin. Thanks are due to Mr Fukunaga, Mr Suzuki and Ms. Junko Nagai, Dai Nippon Printing, Mr Werner Breitfuß, University of Tokyo, and Prof. Helmut Prendinger, Prof. Nubuhiro Fuyurama and his research group, Assistant Prof. Mayumi Bono at the National Institute of Informatics at Tokyo, Ms. Kumiko Campbell, Mr. Tsuyoshi Okita, Ms. Sally Marshall and Mr. Naoto Nishio. We would also like to thank the LiLT reviewers and, last but not least, the anonymous participants of the study.

References

André, E. and T. Rist. 1993. The design of illustrated documents as a planning task. In M. Maybury, ed., *Intelligent Multimedia Interfaces*, pages 94–116. AAAI Press.

André, E., T. Rist, S. Mulken, and M. Van Klesen. 2000. The automated design of believable dialogues for animated presentation teams. In *Embodied Conversational Agents*, pages 220–255. MIT Press.

Arts, A., A. Maes, L. Noordman, and C. Jansen. 2010. Overspecification facilitates object identification. *Journal of Pragmatics* 43(1):361–374.

Arts, A., A. Maes, L. Noordman, and C. Jansen. To appear. Overspecification in written instruction. *Linguistics* .

Bangerter, A. 2004. Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science* 15:415–419.

Belke, E. and A. Meyer. 2002. Tracking the time course of multidimensional stimulus discrimination: Analysis of viewing patterns and processing times during same-different decisions. *European Journal Cognitive Psychology* 14(2):237–266.

Belz, A. 1994. That's nice ... what can you do with it? *Computational Linguistics* 35(1):111–118.

Belz, A. and A. Gatt. 2008. Intrinsic vs. extrinsic evaluation measures for referring expression generation. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL'08)*.

Belz, A. and E. Kow. 2010. The GREC challenges 2010: overview and evaluation results. In *Proceedings of the 6th International Natural Language Generation Conference*, pages 219–229. Association for Computational Linguistics.

Belz, A., E. Kow, J. Viethen, and A. Gatt. 2010. Generating referring expressions in context: The GREC task evaluation challenges. In *Empirical Methods in Natural Language Generation*, vol. 5980 of *Lecture Notes in Computer Science*, pages 294–327. Springer.

Belz, A. and E. Reiter. 2009. An investigation into the validity of some metrics for automatically evaluating NLG systems. *Computational Linguistics* 35(4):529–558.

Breitfuss, W., I. Van der Sluis, S. Luz, H. Prendinger, and M. Ishizuka. 2009. Evaluating an algorithm for the generation of multimodal referring expressions in a virtual world: A pilot study. In *Proceedings of the International Conference on Intelligent Virtual Agents (IVA-09)*. 2009.

Bühler, K. 1934. *Sprachtheorie: Die Darstellungsfunktion der Sprache*. Fischer, Jena.

Byron, D. and L. Stoia. 2005. An analysis of proximity markers in collaborative dialogs. *Proceedings from the Annual Meeting of the Chicago Linguistic Society* 41(2):17–32.

Callaway, C. and J. Lester. 2002. Pronominalization in generated discourse and dialogue. In *Proceedings of of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02)*.

Carletta, J., A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics* 23:13–31.

Claassen, W. 1992. Generating referring expressions in a multimodal environment. In R. Dale, E. Hovy, D. Rösner, and O. Stock, eds., *Aspects of Automated Natural Language Generation*, vol. 587 of *LNCS*, pages 263–276. Springer.

Clark, H. 1996. *Using Language*. Cambridge University Press.

Clark, H. and A. Bangerter. 2004. Changing ideas about reference. In I. Noveck and D. Sperber, eds., *Experimental Pragmatics*, pages 25–49. Palgrave Macmillan, New York.

Clark, H. and D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition* 22:1–39.

Dale, R. 1992. *Generating referring expressions: Constructing descriptions in a domain of objects and processes*. Bradford Books, MIT Press, Cambridge, Massachusetts.

Dale, R. and E. Reiter. 1995. Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science* 18:233–263.

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

Van Deemter, K. 2006. Generating referring expressions that involve gradable properties. *Computational Linguistics* 32(2):195–222.

Van Deemter, K. and A. Gatt. 2009. Beyond DICE: Measuring the quality of a referring expression. In *Proceedings of the Workshop on Production of Referring Expressions: Bridging Computational and Psycholinguistic Approaches (preCogSci-09)*.

Van Deemter, K., A. Gatt, I. Van der Sluis, and R. Power. To Appear. Generation of referring expressions: Assessing the incremental algorithm. *Cognitive Science* .

Van Deemter, K. and E. Krahmer. 2006. Graphs and booleans: On the generation of referring expressions. In H. Bunt and R. Muskens, eds., *Computing Meaning, Studies in Linguistics and Philosophy*, vol. 3. Kluwer, Dordrecht.

Van Deemter, K., B. Krenn, P. Piwek, M. Klesen, M. Schroeder, and S. Baumann. 2008. Fully generated scripted dialogue for embodied agents. *Artificial Intelligence* 172/10:1219–1244.

Di Eugenio, B., P. Jordan, R. Thomason, and J. Moore. 2000. The agreement process: An empirical investigation of human-human computer-mediated collaborative dialogues. *International Journal on Human-Computer Studies* 6:1017–1076.

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

Ford, W. and D. Olson. 1983. The elaboration of the noun phrase in children's object descriptions. *Journal of Experimental Child Psychology* 19:371–382.

Funakoshi, K., S. Watanabe, N. Kuriyama, and T. Tokunaga. 2004. Generating referring expressions using perceptual groups. In *Proceedings of the third International Conference on Natural Language Generation (INLG-04)*.

Funakoshi, K., S. Watanabe, and T. Tokunaga. 2006. Group-based generation of referring expressions. In *Proceedings of the 4th International Conference on Natural Language Generation (INLG-06)*, pages 73–80.

Gardent, C. 2002. Generating minimal definite descriptions. In *Proceedings of the 40st Annual Meeting of the Association for Computational Linguistics (ACL-02)*.

Gardent, C. and K. Striegnitz. 2007. Generating bridging definite descriptions. In *Computing Meaning*, vol. 3. Springer, Netherlands.

Gatt, A. 2006. Generating collective spatial references. In *Proceedings of the 30st Annual Conference of the Cognitive Science Society (CogSci-06)*.

Gatt, A. and A. Belz. 2010. Introducing shared tasks to NLG: The TUNA shared task evaluation challenges. In *Empirical Methods in Natural Language Generation*, vol. 5980 of *Lecture Notes in Computer Science*. Springer.

Gupta, S. and A. Stent. 2005. Automatic evaluation of referring expression generation using corpora. In *Proceedings of the 1st Workshop on Using Corpora in NLG, Birmingham, UK*.

Hasegawa, S. 2000. Danwa kozou to ko so a. In Y. Kusanagi, ed., *Gendai nihongo no goi, bunpou*. Tokyo: Kuroshio.

Henschel, R., H. Cheng, and M. Poesio. 2000. Pronominalization revisited. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING'00)*.

Horacek, H. 1995. More on generating referring expressions. In *Proceedings of the 5th European Workshop on Natural Language Generation (EWNLG'95)*, pages 43–58.

Iida, M. 1997. Discourse coherence and shifting centers in Japanese texts. In E. Prince, A. Joshi, and L. Walker, eds., *Centering in Discourse*. Oxford University Press.

Janarthanam, S. and O. Lemon. 2010. Learning to adapt to unknown users: Referring expression generation in spoken dialogue systems. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL 2010)*.

Jordan, P. 2000. *Intentional Influences on Object Redescriptions in Dialogue: Evidence from an Empirical Study*. Ph.D. thesis, Intelligent Systems Program, University of Pittsburgh.

Jordan, P. and M. Walker. 2005. Learning content selection rules for generating object descriptions in dialogue. *Journal of Artificial Intelligence Research* 24:157–194.

Kameyama, M. 1997. Intrasentential Centering. In E. Prince, A. Joshi, and L. Walker, eds., *Centering in Discourse*. Oxford University Press.

Kelleher, J., F. Costello, and J. Van Genabith. 2005. Dynamically structuring, updating and interrelating representations of visual and linguistic discourse context. In E. Reiter and D. Roy, eds., *Special Issue of Artificial Intelligence Journal on Connecting Language to the World*, vol. 167. Elsevier,.

Kelleher, J. and G. Kruijff. 2006. Incremental generation of spatial referring expressions in situated dialog. In *In Proceedings of the joint 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, (ACL/COLING-06)*.

Koolen, R., A. Gatt, M. Goudbeek, and E. Krahmer. 2009. Need I say more? on factors causing referential overspecification. In *Proceedings of the Workshop on Production of Referring Expressions: Bridging Computational and Psycholinguistic Approaches (preCogSci-09)*.

Krahmer, E. and M. Theune. 2002. Efficient context-sensitive generation of referring expressions. In K. van Deemter and R. Kibble, eds., *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*. CSLI Publications.

Krahmer, E., S van Erk, and A. Verleg. 2003. Graph-based generation of referring expressions. *Computational Linguistics* 29(1):53–72.

Kranstedt, A., A. Lücking, T. Pfeiffer, H. Rieser, and I. Wachsmuth. 2006. Deictic object reference in task-oriented dialogue. In G. Rickheit and I. Wachsmuth, eds., *Situated Communication*. Mouton de Gruyter.

Lester, J., J. Voerman, S. Towns, and C. Callaway. 1997. Deictic believability: Coordinating gesture, locomotion and speech in lifelike pedagogical agents. *Applied Artificial Intelligence* 13(4-5):383–414.

De Lucena, D., B. Pereira, and I. Paraboni. 2010. From semantic properties to surface text: The generation of domain object descriptions. *Inteligencia Artificial* 14(45):48–58.

McCoy, K. and M. Strube. 1999. Taking time to structure discourse: Pronoun generation beyond accessibility. In *Proceedings of the 23st Annual Conference of the Cognitive Science Society (CogSci-99)*.

Morita, Y. 2002. *Nihongo bunpou no hassou*, chap. Shijigo no imi to bunpou: Ko so a no shosou. Tokyo: Hitsuji Shobou.

Pechmann, T. 1989. Incremental speech production and referential overspecification. *Linguistics* 27:89–110.

Piwek, P. 2008. Presenting arguments as fictive dialogue. In *Proceedings of the Workshop on Computational Models of Natural Argument (CMNA-08)*.

Piwek, P., R. Beun, and A. Cremers. 2008. 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics* 40:694–718.

Piwek, P. and A. Cremers. 1996. Dutch and English demonstratives: A comparison. *Language Sciences* 18(3-4):835–851.

Poesio, M., R. Stevenson, B. Di Eugenio, and J. Hitzeman. 2004. Centering: A parametric theory and its instantiations. *Computational Linguistics* 30(3):309–363.

Reithinger, N. 1992. The performance of an incremental generation component for multi-modal dialog contributions. In R. Dale, E. Hovy, D. Rösner, and O. Stock, eds., *Aspects of Automated Natural Language Generation*, vol. 587 of *LNCS*, pages 263–276. Springer.

Rist, T., E. André, and S. Baldes. 2003. A flexible platform for building applications with life-like characters. In *Proc. of the International Conference on Intelligent User Interfaces*, pages 158–165.

Siddharthan, A. and A. Copestake. 2004. Generating referring expressions in open domains. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04)*.

Van der Sluis, I. and E. Krahmer. 2004. The influence of target size and distance on the production of speech and gesture in multimodal referring expressions. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP-04)*. Jeju, Korea.

Van der Sluis, I. and E. Krahmer. 2007. Generating multimodal referring expressions. *Discourse Processes* 44(3):145–174.

Van der Sluis, I., S. Luz W. Breitfuß, M. Ishizuka, and H. Prendinger. To appear. Cross-cultural assessment of automatically generated multimodal referring expressions in a virtual world. *International Journal of Human Computer Studies* .

Spanger, P., Y. Masaaki, I. Ryu, and T. Takenobu. 2009. A Japanese corpus of referring expressions used in a situated collaboration task. In *Proceedings of the 12th European Workshop on Natural Language Generation (ENLG-09)*.

Tanaka, S. and Y. Matumoto. 1997. *Kukan to Idou no Hyogen*. Tokyo: Kenkyusya.

Tenbrink, T. 2004. Identifying objects on the basis of spatial contrast: An empirical study. In *Spatial Cognition IV. Reasoning, Action, and Interaction: International Conference Spatial Cognition 2004*. Frauenchiemsee, Germany: Springer-Verlag GmbH.

Theune, M., R. Koolen, and E. Krahmer. 2010. Cross-linguistic attribute selection for reg: Comparing dutch and english. In *Proceedings of the 6th International Natural Language Generation Conference (INLG 2010)*, pages 191–196.

Thorisson, K. 1994. Simulated perceptual grouping: An application to human computer interaction. In *Proceedings of the 18st Annual Conference of the Cognitive Science Society (CogSci-94)*, pages 876–881. Atlanta.

Tokunaga, T., T. Koyama, and S. Saito. 2005. Meaning of Japanese spatial nouns. In *Proceedings of the ACL-SIGSEM Workshop on The Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*.

Viethen, J. and R. Dale. 2006. Algorithms for generating referring expressions: Do they do what people do? In *Proceedings of the 4th International Conference on Natural Language Generation (INLG-04)*, pages 63–70.

Walker, M., M. Iida, and S. Cote. 1994. Japanese Discourse and the Process of Centering. *Computational Linguistics* 20(2):193–232.

Whitehurst, G. and S. Sonnenschein. 1978. The development of communication: Attribute variation leads to contrast failure. *J. of Experim. Child Psychology* 25:490–504.

Williams, S., P. Piwek, and R. Power. 2007. Generating monologue and dialogue to present personalised medical information to patients. In *Proceedings of the 11th European Workshop on Natural Language Generation(ENLG-07)*.