Producing Referring Expressions in Dialogue: Insights from a Translation Exercise

Ielka van der Sluis (ielka.vandersluis@cs.tcd.ie)
Department of Computer Science
Trinity College Dublin, Ireland

Junko Nagai (jnagai@y8.dion.ne.jp)
Research Consultant
Dai Nippon Printing, Tokyo

Saturnino Luz (luzs@cs.tcd.ie)
Department of Computer Science
Trinity College Dublin, Ireland

Abstract

This paper discusses some of the challenges which arise in comparing the outputs of an algorithm for generating referring expressions across languages and cultures. The context in which the algorithm is employed and evaluated is a virtual environment, and the referring expressions in question form part of a dialogue script "acted out" by two virtual agents in a furniture shop. This setup was created in order to enable us to assess perceptions of a scene by English- and Japanese-speaking subjects with respect to naturalness of dialogue and behaviour of virtual agents, among other things. We show that the process of translating the dialogues from English into Japanese reveals a variety of contextual factors which need to be taken into account for the generation and evaluation of dialogues to be successful in the target language. We focus on issues related to the use of referring expressions, specially: the utility of different types of attributes in the identification of objects, the realisation of locative expressions, and how to deal with the absence of a distinction between singulars and plurals in Japanese. These issues impact on the design and evaluation of algorithms for generating referring expressions in interactive situations, and call into question the extent to which current algorithms are transferrable between languages and cultures,

Keywords: generation of referring expressions, realisation of referring expressions, translation, cross-cultural differences.

Introduction

The generation of referring expressions (GRE) is a central task in Natural Language Generation (NLG), and various algorithms which automatically produce referring expressions have been developed. Recent examples include (Gardent, 2002; Jordan & Walker, 2000; Krahmer, Erk, & Verleg, 2003; Van Deemter, 2002, 2006; Van Deemter & Krahmer, 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 (Dale & 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.

Since human communication includes gestures as well as language some GRE research has also focussed on referring expressions that include pointing gestures and various algorithms for the generation of multimodal referring expressions have been proposed (cf. André & Rist, 1993; Claassen, 1992; Kranstedt, Lücking, Pfeiffer, Rieser, & Wachsmuth, 2006; Lester, Voerman, Towns, & Callaway, 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. The algorithm by Van der Sluis and Krahmer approaches GRE as a compositional task in which language and gestures are combined in a natural way and in which a pointing gesture does not always need to be precise. The algorithm co-relates speech and gesture with respect to the distance between the target referent and the pointing device (in this paper, this would be the finger or hand of the virtual character). The decision to point is based on a notion of effort for which the algorithm uses a cost function.

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 pointing gesture and the use of limited linguistic information (e.g., 'this one'). Alternatively, the algorithm could generate a less precise pointing gesture including other objects in its scope. In this case, more linguistic information has to be added to the referring expression 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 on the side of a room opposite to where the agent stands. For a detailed description of the algorithm see (Van der Sluis & Krahmer, 2007).

The work presented in this paper is part of a project which investigates human perception of the outputs generated by the above mentioned algorithm in a virtual world among subjects of different cultural backgrounds. The study involves subjects in Dublin and in Tokyo who are asked to judge three versions of a dialogue which only differ in the kind of multimodal referring expressions used. The dialogues were originally written in English and subsequently translated into Japanese for the study. The goal of the translation was to produce a Japanese dialogue in which the referring expressions 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 variables other than the choice of referring expression generation might have in the results of the experiment.

This translation process revealed various issues about the use of referring expressions which could impact on the subjects' perceptions of the output of the GRE algorithm. These issues related to the utility of attributes for object identification, the realisation of locative expressions, and the absence of a distinction between singulars and plurals 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 real 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, as also discussed by Van Deemter and Gatt (2009), in this volume.

The paper is structured as follows. A description of the virtual environment and dialogue used in the study is presented in order to better contextualise our observations on linguistic referring expressions. This description is accompanied by the preliminary comments and requests for clarification made by the Japanese translator in preparation for the task. This is followed by an analysis of the issues which arose in the process of translating referring expressions from English into Japanese for the purposes of the study outlined above. Finally, the implications for GRE research are discussed.

The Setting

As mentioned above, the referring expressions discussed in this paper were generated as part of a study in which subjects watched the unfolding of a scripted dialogue situation involving two virtual agents in a furniture shop. The situation was set in *Second Life* (SL), a virtual 3D world accessible over the Internet. The SL environment 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 referring expressions.

Figure 1 illustrates the layout of the furniture shop marking the positions of the furniture items and the agents. Apart from a number of distractor objects, marked as 'other item', the layout shows a number of furniture items that are used for assessing multimodal GRE output: (1) a large red chair (at the bottom left); (2) a large blue desk (at the top left), (3) a small blue desk (next to the large one); (4) a set of large red chairs (in the middle), and (5) a set of small green chairs (next to the set of reds). The agent in the role of furniture seller is able to move close to these items and refer to them using a precise pointing gesture. Alternatively, the agent can stay stationary at the position indicated in the Figure and point in the direction where the target item(s) can be found. The agent in the role of buyer can follow the seller around in the shop. A more detailed description of the environment and a pilot study based on it can be found in (Breitfuss, Van der Sluis, Luz, Prendinger, & Ishizuka, 2009).

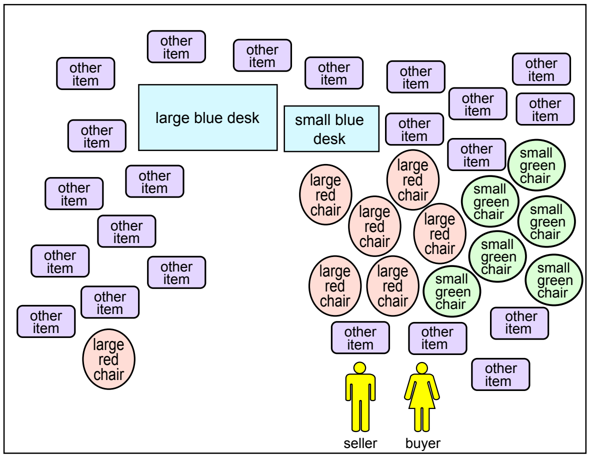

Figure 1: Bird's-eye view of the virtual furniture shop.

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, the virtual furniture shop, its standing and size is correlated with the attitude of the people that populate it and on the phrases and words they use. In the design of the agents and the shop, which was built for a perception study in Dublin, the gender, age and social standing of the agents were not very well defined. The agents had, in the words of the translator, "a sort of abstract countenance that seems to belong to anywhere and thus nowhere". 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. It was decided to use more localised agents in the Japanese as well as in the Dublin set up and to approximate their age at about 25-35 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.

With respect to the graphical features of the set up, the agents were able to produce pointing gestures as well as non-deictic gestures. The non-deictic gestures, like beats and more general body movements, were included to increase the naturalness of the agents. It was observed that the non-deictic gestures displayed by the agents in the English dialogue script, were rather "large" (i.e. occupying much space) compared to the non-deictic gestures observed in a similar scenario in Japan. In the Japanese version, the size of the non-deictic gestures was adapted, so as not to distract the viewer from assessing the referring expressions. As regards pointing gestures, it was recognized that in situations where people are expected to act politely (people working in shops, hotels and other sectors of the service industry are expected to address their customers politely) are trained to produce pointing gestures using their whole hand, while Europeans are more likely to point with just their index finger. In addition, the speed of the gestural movements is somehow age dependent (e.g., easy pace for the elderly, somewhat quicker for the younger generation). Because the target audience for the study is around 20-30 years, the gesture speed was left as it was.

The virtual furniture shop that was built for the study displayed the furniture in a relatively large space, which made it easy for the agents to move around. However, in Japan people are quite often used to small shops that have their merchandise crammed into a small space. Space and distance that is felt to be natural, comfortable or usual by many Japanese people may be smaller and more stuffed compared to what is comfortable for Europeans. However, to avoid situations in which the agents would stumble over furniture items while moving around in the shop, it was decided to keep the space for the agents to move around.

Dialogue

We employed a Fully Generated Scripted Dialogue (FGSD) approach (André, Rist, Mulken, & van Klesen, 2000; Williams, Piwek, & Power, 2007) to evaluate the output of the Van der Sluis and 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 the purpose of assessing the perception of the output of the Van der Sluis and Krahmer algorithm, a dialogue script was written for two virtual characters in a virtual furniture store. The furniture shop contains over 43 objects, 13 of which are actually referred to in the dialogues. The other items in the shop are 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, Jordan, Thomason, & Moore, 2000) and the TUNA corpus (Gatt, Van der Sluis, & Van Deemter, 2007), and we hoped that these knowledge sources would help us to construct a believable dialogue.

The dialogue consists of 19 utterances with 5 first mention references to furniture items (3 singletons and 2 sets) 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. The dialogue was used as a template in which the 5 referring expressions were varied (3 referring to singular and 2 referring to sets of items). The referring expressions used to fill out these slots can be automatically reproduced with the Van der Sluis and Kramer algorithm. The output of this algorithm is dependent on the cost function it uses and varies in the amount of linguistic information included in the description and in the type of pointing gesture to be produced by the agent (ie. changing the distance between the agent and the target object).

Three different versions of the dialogue script were written so that only the five referring expressions differed. Three types of output were implemented in three dialogues, with referring expressions ranging over two extremes with respect to linguistic and pointing information. One extreme, the imprecise version, used a version of the algorithm that generates very detailed linguistic descriptions of objects in which all the attributes of the target object were included. The pointing gestures generated to accompany these descriptions are, however, vague and the virtual agent can direct them from a considerable distance from the target object. In this version the agents are stationary in the positions indicated in Figure 1. The other extreme, the precise version, used another version of the algorithm that generates limited linguistic information (e.g. 'this one') combined with precise pointing gestures. In this version, the agents move through the shop along the furniture items under discussion. Between these two extremes a 'mixed version'. was implemented, in which 2 targets in the dialogue were identified with precise pointing gestures (1 singleton and 1 set) and 3 targets were identified with imprecise pointing gestures (2 singletons and 1 set).

Results from a pilot study, used for validation of the dialogues, showed that the dialogues were acceptable for an Irish audience. For more details about the set up of our study we refer to (Breitfuss et al., 2009).

GRE: From English to Japanese

The goal of the translation was to produce a Japanese dialogue in which the referring expressions 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 reason, the beginning of the dialogue was altered considerably in the Japanese version, both in speech and in gestures. 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 pushy or bold. In general, the English dialogue was too verbose compared to a 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.

This paper focuses on one of the three dialogues used to validate the Van der Sluis and Krahmer algorithm, namely the one in which each of the five distinguishing referring expressions contain all attributes that are known to the algorithm about the objects (ie. 'type', 'colour', 'size', 'location'). The dialogue script is implemented so that the furniture seller produces the linguistic descriptions in combination with deictic gestures that point in the direction of the target objects, which are located at a considerable distance from where the two agents stand. The location of the agents and furniture is similar to the layout depicted in Figure 1. The remainder of this paper only addresses the linguistic parts of the multimodal referring expressions shown in Table 1.

In translating the referring expressions from English to Japanese, a number of issues arose, such as the fact that the definite article, anaphoric expressions like 'one' or 'the ones', and pronouns like '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 also the Japanese 'sore' often replaces the English 'it', this does not mean the words 'mono' and 'sore' have the same functions as the English pronouns, 'one', 'ones' and 'it'. 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 referring expression to its antecedent.

When looking at the translations in Table 1, 'one', in expression (1), was translated with 'isu' meaning 'chair'. The pronoun 'it' in expression (3), 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, which is emphasised by the use of 'sono', which in this case means 'none other than (what was mentioned earlier)'. In expression (5) 'the red ones' was not omitted, because 'the red ones' 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 referring expression (e.g. the small green chairs). Also, because the Japanese language lacks the morphological means to indicate plurality, the translator has 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. In what follows the choice of attributes, the realisation of 'location' and cardinality is addressed in more detail.

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 & Olson, 1983; Whitehurst & Sonnenschein, 1978; Pechmann, 1989; Belke & 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. 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 referring expression. For instance, description (1), in Table 1, was embedded in the dialogue as follows:

> 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.'

In translating English referring expressions 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 'naturalness' of the dialogue, which would obviously be a problem for a cross cultural study where naturalness is one of the variables being studied. One way to settle the first issue would be through corpus-based study (cf. Spanger, Masaaki, Ryu, & Takenobu, 2009). 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 GRE research.

Dimensions of 'location'

Locative expressions are generally used in referring expressions 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 (1), (2) and (4) include an absolute locative expression, while descriptions (3) and (5) include a relative locative expression. Translation of these English locative expressions into Japanese was not easy. To begin with, there is no straightforward correspondence between English prepositions and Japanese postpositional particles. In English, spatial relations are often

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

(1) *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.

(2) *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

(3) *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)

(4) *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/chairs

Retranslation: large red chair/chairs in not too far or too near place/direction in middle

(5) *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/chairs		next		small	green-colour		chair/chairs

Retranslation: small green-colour chair/chairs next to red chair/chairs

represented by prepositions (e.g., ’in’, ’above’), whereas in Japanese spatial relations are often represented by spatial nouns and postpositional particles, or by postpositional particles alone (Tokunaga, Koyama, & Saito, 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 in terms of space, time and emotion, the relationship between the addressee and speaker, in terms of whether they both share the same spatial or temporal or emotional perspective, also 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 timeframe, 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 (1) uses a demonstrative pronoun from the ’ko’ family (‘kochira’) which indicates nearness. Expression (2) uses a demonstrative pronoun of the ’a’ family (‘achira’) which indicates a ‘large distance’. Expression (3) 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 (4) 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 (5) 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 referring expression. In this way 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 postpositional 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 postpositional particle 'no'. In this case, the choice between 'ko' and 'so' words would depend on the speaker's judgment 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 (1), 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 shop keeper speaking in socially polite form, would choose a polite form of demonstrative to convey his message to his addressee, the furniture buyer.

With respect to 'location' the work on GRE has mainly focussed on the perceptual grouping of objects (Funakoshi, Watanabe, Kuriyama, & Tokunaga, 2004; Funakoshi, Watanabe, & Tokunaga, 2006; Gatt, 2006; Thorisson, 1994), that is on spatial information only. In addition, there is research on the use of demonstratives (Piwek & Cremers, 1996). To our knowledge, distance and dimensions of time and emotion as they can be indicated with the Japanese 'a', 'so' and 'ko' system have not been addressed.

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 (1) (ie. '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. 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 & Matumoto, 1997). Thus, a combined expression of 'definite article' and 'absolute location marker' in the English language is transformed into a combined expression of 'speaker's-perception-dependent demonstrative of the ko family' and '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 ...')[1]. 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 2, 'in the back' ('achira no oku ...'), and in expression 4, 'in the middle' ('sochira no nakahodo ...'), was handled in a similar fashion.

With regard to relative locative expressions, a slight discrepancy of meaning was detected between the seemingly equivalent English-Japanese 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, 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 & Reiter, 1995; Krahmer & Theune, 2002). Implemention in this particular furniture shop, renders 'next to' for both descriptions (3) and (5), because there is no other desk located to the left of the object referred to by the pronoun 'it' (description (3)), and there are no chairs to the left of the 'the red ones' (description (5)). Now, what would producing a basic level value such as 'next to' do for the Japanese addressee? Arguably, it would make the referring expression 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 the producing 'next to' makes interpretation 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'.

Singulars versus Plurals

As illustrated by Van Deemter and Krahmer (2006) a version of the graph-based algorithm that can generate referring expressions for sets of objects is straightforward. Hence, we

[1] 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

decided to include references to sets of objects (descriptions (4) and (5)) in this perception study. However, in the Japanese language nouns do not have a plural form. The singular chair of description (1), and the set of chairs referred to in description (4), 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 need to count the objects in a group first before uttering the referring expression. Similarly, the furniture buyer would need to count the objects in a group before she could be sure to have identified the correct set.

Although the Japanese translation lacks the information that comes from the morphology of the English noun, the translations of descriptions (4) and (5) are still distinguishing in the setting because of the inclusion of location. However, in a sense, the lack of cardinal information makes the Japanese referring expression less redundant than the English one. But how to handle cases in which the cardinal information is necessary for distinguishing objects? A possible way to add a sense of plurality to the Japanese translations of descriptions (4) and (5) 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.

For the sake of the cross cultural study, one might question if the same algorithmic output is tested when a part of the distinguishing features in the referring expressions is omitted. With respect to GRE research, there may be a need to introduce more distinctions in the knowledge representation used in GRE. So far the work in GRE has not paid much attention to knowledge representation and it seems that GRE algorithms make some implicit assumptions about the target language. In addition, with respect to GRE evaluation purposes, it might be that the current use of similarity metrics is too shallow to point out the differences between bi-lingual pairs of referring expressions (cf. Van Deemter & Gatt, 2009).

Conclusions

This paper described issues which were dealt with in the process of translating English object descriptions into the Japanese language. The referring expressions under consideration are part of a dialogue script written to assess the output of the multimodal GRE algorithm proposed by Van der Sluis and Krahmer (2007). The dialogue is 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 automatically produced referring expressions as similar as possible to the English originals. Our detailed analysis of the challenges that we came accross in the translation process reveals a number of issues relevant to the work in GRE, especially the work that targets life-like contexts.

In the translation process a number of differences between the English and Japanese language were obvious from the start. For instance, it was pointed out that definite articles (e.g. 'the'), anaphoric expressions like 'one' or 'ones' and pronouns like 'it' do not have an exact equivalent in the Japanese language. Neither does the Japanese language know plural noun forms. In the translations these meanings had to be captured by other means or omitted completely.

In contrast to English, the Japanese language has a rich vocabular to express distance in terms of, not only space, but also in terms of time and emotion. These dimensions seem to be of particular importance when using referring expression in a specific context or discourse. It was also found that the use of specific and basic level values in locative expressions differs between English and Japanese. This difference may arise from cross-cultural differences between the 'sense of physical closeness/distance'.

It also appeared that the relevance or salience of the attributes used in a description 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 accross languages. Obviously translation of the referring expressions might in itself affect the 'naturalness' of the dialogue. To avoid this bias in the referring expressions, cross-cultural GRE research needs corpora in languages other than English (cf. Koolen, Gatt, Goudbeek, & Krahmer, 2009; Spanger et al., 2009).

The above findings raise the issue of how transferrable 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 generate referring expressions.

Acknowledgments

This research is funded by Science Foundation Ireland as part of the CNGL project (www.cngl.ie). Thanks are due to Mr Fukunaga and Mr Suzuki, at Dai Nippon Printing, Mr Werner Breitfuß, at the University of Tokyo, and Prof. Helmut Prendinger, at the National Institute of Informatics.

References

André, E., & Rist, T. (1993). The design of illustrated documents as a planning task. In M. Maybury (Ed.), *Intelligent multimedia interfaces* (p. 94-116). AAAI Press.

André, E., Rist, T., Mulken, S., & van Klesen, M. (2000). The automated design of believable dialogues for animated presentation teams. In *Embodied conversational agents.* MIT Press.

Belke, E., & Meyer, A. (2002). Tracking the time course of multidimensional stimulus discrimination: Analysis of viewing patterns and processing times during same-different decisions. *Eur. J. Cogn. Psychol.*, *14*(2), 237-266.

Breitfuss, W., Van der Sluis, I., Luz, S., Prendinger, H., & Ishizuka, M. (2009). Evaluating an algorithm for the generation of multimodal referring expressions in a virtual world: a pilot study. In *Proc. of IVA-09.* (To appear.)

Claassen, W. (1992). Generating referring expressions in a multimodal environment. In R. Dale, E. Hovy, D. Rösner, & O. Stock (Eds.), *Aspects of automated natural language generation* (Vol. 587, p. 263-276). Springer.

Dale, R., & Reiter, E. (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., & Gatt, A. (2009). Beyond dice: measuring the quality of a referring expression. In *Proc. of the workshop preCogSci-09.*

Van Deemter, K., & Krahmer, E. (2006). Graphs and booleans: on the generation of referring expressions. In H. Bunt & R. Muskens (Eds.), *Computing meaning, Studies in linguistics and philosophy* (Vol. 3). Kluwer, Dordrecht.

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

Di Eugenio, B., Jordan, P., Thomason, R., & Moore, J. (2000). The agreement process: an empirical investigation of human-human computer-mediated collaborative dialogues. *Intl. Journ. Human-Comp. Studies*, *6*, 1017-1076.

Ford, W., & Olson, D. (1983). The elaboration of the noun phrase in children's object descriptions. *Journal of Experimental Child Psychology*, *19*, 371-382.

Funakoshi, K., Watanabe, S., Kuriyama, N., & Tokunaga, T. (2004). Generating referring expressions using perceptual groups. In *Proc. of the INLG-04.*

Funakoshi, K., Watanabe, S., & Tokunaga, T. (2006). Group-based generation of referring expressions. In *Proc. of the INLG-06* (p. 73-80).

Gardent, C. (2002). Generating minimal definite descriptions. In *Proc. of the ACL-02.*

Gatt, A. (2006). Generating collective spatial references. In *Proc. of CogSci-06.*

Gatt, A., Van der Sluis, I., & Van Deemter, K. (2007). Evaluating algorithms for the generation of referring expressions using a balanced corpus. In *Proc. of the ENLG-07.*

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

Jordan, P., & Walker, M. (2000). Learning attribute selections for non-pronominal expressions. In *Proc. of the ACL-00.*

Koolen, R., Gatt, A., Goudbeek, M., & Krahmer, E. (2009). Need i say more? on factors causing referential overspecification. In *Proc. of the workshop preCogSci-09.*

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

Krahmer, E., & Theune, M. (2002). Efficient context-sensitive generation of referring expressions. In K. van Deemter & R. Kibble (Eds.), *Information sharing: Reference and presupposition in language generation and interpretation.* CSLI Publications.

Kranstedt, A., Lücking, A., Pfeiffer, T., Rieser, H., & Wachsmuth, I. (2006). Deictic object reference in task-oriented dialogue. In G. Rickheit & I. Wachsmuth (Eds.), *Situated communication.* Mouton de Gruyter.

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

Morita, Y. (2002). *Nihongo bunpou no hassou.* In (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 *Proc. of the CMNA-08.*

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

Reithinger, N. (1992). The performance of an incremental generation component for multi-modal dialog contributions. In R. Dale, E. Hovy, D. Rösner, & O. Stock (Eds.), *Aspects of automated natural language generation* (Vol. 587, p. 263-276). Springer.

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

Spanger, P., Masaaki, Y., Ryu, I., & Takenobu, T. (2009). A Japanese corpus of referring expressions used in a situated collaboration task. In *Proc. of the ENLG-09.*

Tanaka, S., & Matumoto, Y. (1997). *Kukan to idou no hyogen.* Tokyo: Kenkyusya.

Thorisson, K. (1994). Simulated perceptual grouping: An application to human computer interaction. In *Proc. of CogSci-94* (p. 876-881). Atlanta.

Tokunaga, T., Koyama, T., & Saito, S. (2005). Meaning of Japanese spatial nouns. In *Proc. ACL-SIGSEM workshop on the linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications.*

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

Williams, S., Piwek, P., & Power, R. (2007). Generating monologue and dialogue to present personalised medical information to patients. In *Proc. of ENLG-07.*