

The Interplay between Joint Attention, Physical Proximity, and Pointing Gesture in Demonstrative Choice

David Peeters (david.peeters@mpi.nl)

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands
Radboud University, Donders Institute for Brain, Cognition, and Behaviour, Nijmegen, The Netherlands

Zeynep Azar (zeynep.azar@mpi.nl)

Radboud University, Center for Language Studies, Nijmegen, The Netherlands

Aslı Özyürek (asli.ozyurek@mpi.nl)

Radboud University, Center for Language Studies, Nijmegen, The Netherlands
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Abstract

A fundamental property of language is that it allows us to establish triadic joint attention to a referent, for instance by the use of spatial demonstratives. Traditional accounts of demonstrative choice focused on the physical proximity of the referent to the interlocutors. However, recent work taking into account the multimodal context in which spatial demonstrative use is generally embedded shows that such accounts are too simplistic. Using a controlled elicitation task, we here tested the differential roles of visual joint attention, physical proximity of a referent, and use of a pointing gesture in demonstrative choice in Dutch. It was found that ‘proximal’ demonstratives were used in a speaker-anchored way to refer to objects nearby the speaker. ‘Distal’ demonstratives were used for referents not nearby the speaker, but also in an addressee-anchored way, i.e. when the referent was in the addressee’s focus of visual attention. Pointing gestures were closely tied to demonstratives but not to the use of (in)definite articles. These findings show that demonstrative choice is dependent on a subtle interplay between different context-dependent factors as a function of both speaker- and addressee-anchored perspectives. Findings are discussed in terms of demonstrative systems and multimodal reference production in general.

Keywords: Spatial Demonstratives; Deixis; Joint Attention; Reference; Pointing Gesture

Introduction

Establishing triadic joint attention to a referent is a very basic human communicative ability (Carpenter, Nagell, & Tomasello, 1998) and a common way to do so is by using a linguistic referring expression, often combined with a manual pointing gesture and a shift of gaze (Bakeman & Adamson, 1984; Clark & Bangerter, 2004; Diessel, 1999). In everyday spoken interaction, people often refer to objects in the extra-linguistic physical surroundings of a conversation (i.e. exophorically) by using demonstrative pronouns and determiners (henceforth: demonstratives) such as *this* and *that* (in English) in their speech. Most languages contain more than one demonstrative type (Diessel, 2005), which implies that people have to choose among different

options when using a demonstrative. Traditional accounts of demonstrative reference are spatialist in nature. Generally they take the relative proximity of a referent to the speaker as a fundamental criterion in demonstrative choice (e.g., Anderson & Keenan, 1985; Halliday & Hasan, 1977; Lyons, 1977, *inter alia*). Such proximity-based accounts argue that one would select a proximal demonstrative to refer to an object that is physically relatively nearby, and a distal demonstrative to refer to an object that is relatively far away. This speaker-anchored proximity-based view is omnipresent in reference grammars (Diessel, 2005) and recent experimental studies provide partial support (e.g., Coventry et al., 2008; Stevens & Zhang, 2013).

However, more and more evidence accumulates showing that such traditional accounts are too simplistic (e.g., Enfield, 2003; Hanks, 1990; Strauss, 2002). Descriptions of demonstrative systems in different languages from different language families (e.g., Dutch, English, Jahai, Jordanian Arabic, Turkish) suggest that factors related to the locus of visual attention of the addressee and/or the cognitive (rather than physical) accessibility of a referent may play an important role in which demonstrative a speaker chooses (Burenhult, 2003; Jarbou, 2010; Küntay & Özyürek, 2006; Piwek et al., 2008; Stevens & Zhang, 2013). Küntay and Özyürek (2006), for instance, show that the Turkish demonstrative *şu* is primarily used to shift the visual attention of the addressee when she does not (yet) visually attend to the referent, independent of the relative physical distance of that referent. As such, joint attention between speaker and addressee focused on a referent is not only often the desired outcome of exophoric demonstrative use (Diessel, 1999), but the presence or absence of joint attention to a referent at the moment a referring expression is instantiated may also drive the choice for a particular demonstrative over another. In general, being able to monitor and follow the gaze of an interlocutor is indeed a pivotal communicative skill and often a prerequisite for successful communication (e.g., Bakeman & Adamson, 1984).

There are some preliminary indications that joint attention may also play a role in demonstrative choice in Dutch, a language with a two-term demonstrative system. Piwek, Beun, and Cremers (2008) investigated demonstrative choice in Dutch in cases of exophoric reference in a building task. Pairs of participants, consisting of an instructor and a builder, constructed a small building using Lego blocks. Participants' speech recorded during the task was analyzed off-line in terms of their focus of attention on a referent. A referent was coded as in the focus of attention, i.e. cognitively relatively accessible, when it was mentioned in the preceding utterance and/or when it was in an area toward which the speaker had explicitly directed the attention of the addressee already. It was found that participants used proximal demonstratives (*dit, deze* in Dutch) to refer to objects that were not in the focus of attention (low cognitive accessibility) and distal demonstratives (*dat, die*) to objects that were in the focus of attention (high cognitive accessibility). On the basis of their results, Piwek et al. (2008) argue against proximity-based views of demonstrative choice.

The operationalization of focus of attention by Piwek et al. (2008), however, does not differentiate between cognitive and visual foci of attention. In the current study we zoom in on the visual focus of attention of the addressee only. Research in different languages has shown that the visual attention of the addressee may have an effect on demonstrative choice (Küntay & Özyürek, 2006; Stevens & Zhang, 2013). In addition, although Piwek et al. (2008) argue against proximity-based views of demonstrative choice, they did not quantify the physical proximity of the referents in their study. It is therefore unclear whether the physical proximity of referents (to speaker and/or addressee) also influenced demonstrative choice and whether this interacted with the addressee's visual focus of attention. In general, previous observational work has often not enabled investigation of interactions between different variables that influence demonstrative choice in a controlled way.

The current study therefore experimentally contrasted three variables that may influence demonstrative choice. In a controlled elicitation task, participants were presented with different visual scenes that induced the production of referring expressions. First, the influence of the visual focus of attention of the addressee (and as such the joint attention between speaker and addressee to a referent) in the visual scenes was varied. Proximity-based theories do not predict a difference in demonstrative choice based on the attentional focus of the addressee. In contrast, if the findings by Piwek et al. (2008) generalize to situations where the visual focus of attention of the addressee is manipulated, this would predict the use of proximal demonstratives for referents that are not in the addressee's visual focus of attention. In such a situation, a proximal demonstrative in Dutch would indeed allow strong indicating to shift the addressee's attention towards the referent (Piwek et al., 2008). Under such an account, Dutch speakers would use distal demonstratives for

referents that are already in the focus of visual attention of the addressee.

Secondly, in orthogonal contrast to the addressee's focus of attention we varied the location of the referent, and as such its physical distance from the speaker and addressee. Proximity-based accounts predict that proximal demonstratives are used for referents close to the speaker and that the use of distal demonstratives increases with an increase in the relative physical distance of the referent from the speaker (e.g., Anderson & Keenan, 1985; Halliday & Hasan, 1977; Lyons, 1977). In contrast, Piwek et al. (2008) argue that the relative physical distance of a referent does not primarily drive demonstrative choice.

A third and novel factor manipulated here is the presence or absence of a pointing gesture. Proximity-based accounts do not predict an influence of this variable. However, observational research suggests that within a language pointing gestures may be more closely tied to some demonstratives than to others (Küntay & Özyürek, 2006; Senft, 2004). The participants in the study by Piwek et al. (2008) always produced a manual pointing gesture when uttering a proximal demonstrative, but not always when using a distal demonstrative. However, it is unclear whether this finding interacts with or is driven by the relative distance of the referents from the speaker.

Finally, the current study not only focuses on the choice of demonstratives but also looks at noun phrases containing (in)definite articles. Demonstratives can be placed within a wider class of referring expressions (e.g., Ariel, 1988) and it is an open question whether similar factors influence the choice of demonstratives and the choice of a particular article in exophoric reference. It is not unlikely that speakers would use more definite than indefinite articles when objects are nearby and/or when their addressee is already visually attending to a referent. Furthermore, pointing gestures may be paired with demonstratives more than with articles, because one function of demonstratives may be to direct the addressee's gaze to the gesture (Bangertner, 2004). Identification of the factors influencing the choice of referring expression is not only theoretically interesting, but may also inform computational models of reference production (see Van Deemter, Gatt, Van Gompel, & Krahmer, 2012).

Method

Participants

Twenty native speakers of Dutch studying in Nijmegen (13 female; mean age 22.2) participated in return for payment. They had normal or corrected-to-normal vision and no history of language impairment.

Materials

The materials consisted of 64 triplets of still images that contained a speaker, an addressee, and an object. Each triplet consisted of an introductory picture, a target picture,

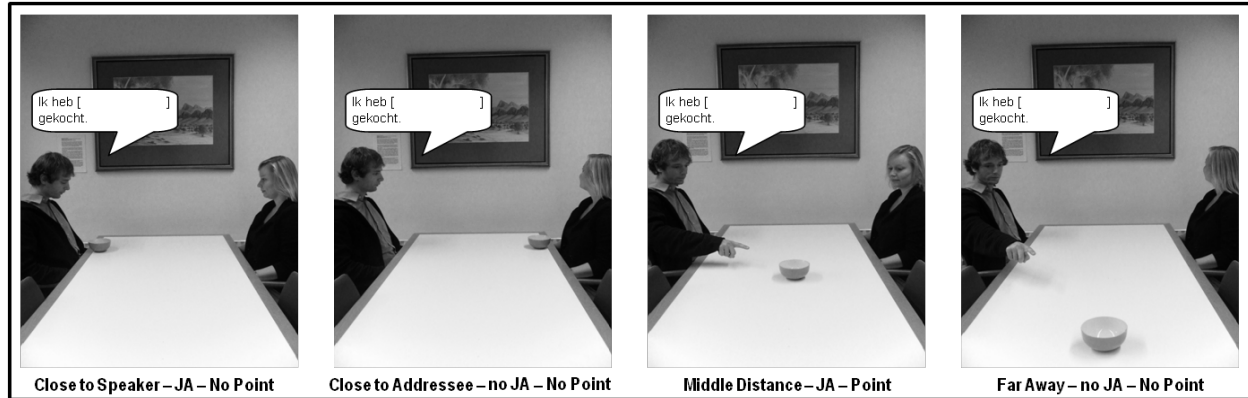


Figure 1: Subset of target pictures used in the elicitation task (converted to grayscale). Below the pictures it is indicated to which condition the picture belonged. The object could be in four different locations. There could be joint attention (JA) or no joint attention (no JA) between speaker and addressee to the object, and the speaker could either make a pointing gesture towards the object or not.

and a concluding picture. Figure 1 shows a subset of the picture materials.

Three independent variables were orthogonally manipulated in the target pictures. First, the location of the object could be close to the speaker, close to the addressee, at middle distance from both speaker and addressee, or relatively far away from speaker and addressee. Second, there could be either visual joint attention or no visual joint attention between speaker and addressee to the object. In the case of no visual joint attention, the speaker looked at the referent object while the addressee looked at another part of the visual scene (e.g., a painting, see Fig.1). Third, the speaker could either produce a pointing gesture towards the object or not. It is not uncommon in natural interactions for speakers to refer to an object using speech and gesture while their addressee is not yet looking at them or their referent (e.g., Küntay & Özyürek, 2006). The three manipulated factors are henceforth called Location, Joint Attention, and Pointing Gesture respectively.

In all target pictures the speaker and addressee were either sitting (n=32) or standing (n=32) opposite each other. The referent object was a ball, a bowl, a bucket, or a cup (n=16 each). Whether the person on the left or the right in the picture was the speaker was counterbalanced and indicated by a text balloon close to the mouth of the speaker. Every target picture contained a text balloon presenting a declarative sentence that was missing a referring expression and could be completed using a demonstrative or an article and a noun. All sentences were of the form ‘Ik heb [] + verb’ (‘I have [] + verb’), which elicited sentences such as ‘Ik heb dit kopje schoongemaakt’ (I have cleaned this cup) or ‘Ik heb die bal meegebracht’ (I have brought that ball). The verb form was specified in the sentences; only the referring expression was left out.

The introductory picture was always the same as the target picture except for the presence of the text balloon, and served to introduce the visual context of each trial. In the

concluding picture there was always joint attention between speaker and addressee to the referent object. As such every referential act successfully resulted in joint attention to the referent.

In addition to the 64 experimental trials, 64 triplets were created that were used as fillers in the task. These images depicted a single speaker and one object. A large number of different objects were used. In these filler trials, the text balloons contained a sentence that would not elicit a demonstrative or article (e.g., ‘What a nice []’).

Procedure

Participants were tested individually in a sound-proof booth. The experiment was presented on a computer screen using *Presentation* software (Neurobehavioral Systems). Participants were instructed to carefully look at the pictures, and read aloud and complete the sentence shown in the picture that was presented in the center of the screen. Trials were presented one by one. Each trial started with an introductory picture at the left part of the screen. After 1500 ms, the target picture appeared in the centre of the screen, while the introductory picture remained visible. After having read aloud and completed the written sentence in the target picture, participants pressed the spacebar on a keyboard, which resulted in presentation of the concluding picture. Together with the other two pictures, this picture remained visible on the screen for another 1500 ms, after which the next trial started. The 128 trials were presented in a unique randomized order for each participant.

Data Coding

Throughout the task, participants’ speech was continuously recorded by a voice recorder (Olympus Imaging Corp.) linked to an external microphone. The elicitation task yielded 128 uttered sentences per participant. Speech was transcribed off-line and each referring expression was coded for the presence of a demonstrative (proximal or distal) or an article (definite or indefinite) preceding the noun. In line

with previous studies (e.g. Piwek et al., 2008), demonstratives were collapsed across grammatical gender in the analysis.

Results

Participants produced a demonstrative or article on 95.3% of all trials. In this dataset, 32.4% of trials contained a demonstrative (9.0% proximal + 23.4% distal) and 67.6% of trials contained an article (20.2% definite + 47.4% indefinite).

Separate repeated measures analyses of variance were carried out on the proportion of use of the proximal and distal demonstrative, and the definite and indefinite article, with the three independent variables (Location, Joint Attention, and Pointing Gesture) as factors (cf. Küntay & Özyürek, 2006). The Greenhouse-Geisser correction was applied in all analyses. Corrected degrees of freedom are reported.

The analysis on the use of proximal demonstratives showed no effect of Joint Attention ($F < 1$). A significant main effect of Location was found, $F(1,22) = 14.62$, $p = .001$, $\eta_p^2 = .435$. Significantly more proximal demonstratives were used when the referent object was close to the speaker compared to the three other referent locations (all p -values $< .01$). In addition, a significant main effect of Pointing was found $F(1,19) = 6.77$, $p = .017$, $\eta_p^2 = .263$. Proximal demonstratives were used significantly more often when the speaker pointed compared to when she did not point.

The analysis on the use of distal demonstratives showed a significant main effect of Joint Attention, $F(1,19) = 5.32$, $p = .033$, $\eta_p^2 = .219$. The proportion of distal demonstratives used was significantly higher in the case of joint attention compared to no joint attention. In addition, a significant main effect of Location was found, $F(2,36) = 12.25$, $p = .001$, $\eta_p^2 = .392$. Participants used significantly fewer demonstratives when the object was close to the speaker compared to the other three locations (all p -values $< .01$). A significant main effect of Pointing, $F(1,19) = 15.05$, $p = .001$, $\eta_p^2 = .442$, reflected that distal demonstratives were used significantly more often when the speaker pointed compared to when she did not point. Finally, a significant Location x Pointing interaction effect was found, $F(2,42) = 3.18$, $p = .047$, $\eta_p^2 = .143$. This interaction effect reflected that the main effect of Pointing only held when the object was close to the addressee ($p = .008$), at middle distance ($p = .013$), or far away ($p = .015$). No main effect of Pointing was found for the object close to the speaker ($p = .716$) when the distal demonstrative was used.

No significant effects were found for the use of definite articles. The analysis of the use of indefinite articles, however, yielded a significant main effect of Pointing, $F(1,19) = 13.84$, $p = .001$, $\eta_p^2 = .421$. Indefinite articles were used significantly more when the speaker did not point compared to when she did point. Figure 2 depicts all results.

Conclusion and Discussion

A fundamental property of language is that it allows us to refer to entities in the world around us (e.g., Clark, 1996; Tomasello, 2008). The current study investigated three factors that might influence how we do this by zooming in on the choice of demonstratives in Dutch. First, we found that distal demonstratives were used more often when there was visual joint attention between speaker and addressee to the referent compared to when there was no joint attention. This finding is in line with the suggestion by Piwek et al. (2008) that the Dutch distal demonstrative is used when no strong indicating is necessary because the referent is already in the focus of attention. Research on other languages has also found an influence of the addressee's attention on demonstrative choice (e.g., Burenhult, 2003; Küntay & Özyürek, 2006). Together, these findings confirm that joint attention may not only be the aim and result of using a referring expression (Diessel, 1999), but also a driving force in demonstrative choice. Interestingly, Dutch differs here from other two-term demonstrative systems (see also Kirsner & Van Heuven, 1988). For instance, in Jordanian Arabic it is the proximal demonstrative that is used for entities with high perceptibility to the addressee (Jarbou, 2010). In contrast, we did not find an influence of the focus of attention of the addressee on the use of the proximal demonstrative in Dutch. Indeed, factors influencing demonstrative choice do not need to be symmetrically expressed in different demonstrative terms within a language (Enfield, 2003).

Second, we found that Dutch speakers also took into account the relative proximity of the referent to the speaker in their demonstrative choice. Referents nearby the speaker elicited proximal demonstratives whereas referents in three physically more distant regions elicited distal demonstratives. No linear increase of distal demonstrative use was found as a function of an increase in relative physical distance from the speaker to the referent. Rather, speakers differentiated between a zone close to the speaker and the rest of the extra-linguistic space. This also explains why the presence of a pointing gesture only influenced the use of a distal demonstrative when the referent was not near the speaker. Arguably, because the distal demonstrative would be used for a less strict demarcated part of space than the proximal demonstrative, a pointing gesture could narrow down the addressee's search space in looking for the intended referent.

In sum, in our study the Dutch proximal demonstrative was used in a speaker-anchored way. Referents nearby the speaker elicited proximal demonstratives and the visual attention of the addressee did not influence the use of this demonstrative. However, Dutch speakers did take into account their addressee in demonstrative choice and the current results suggest that this is marked on the distal demonstrative. Arguably the distal demonstrative may be used both in a speaker-anchored way (for objects not near the speaker) and in an addressee-anchored way, as its use may acknowledge that the referent is already in the

addressee's focus of visual attention. Thus, the attentional state of the addressee and the use of pointing are differentially exploited in the use of different demonstratives. Earlier accounts of demonstrative reference

have sometimes explained demonstrative choice by focusing on a single factor such as proximity (e.g., Anderson & Keenan, 1985; Halliday & Hasan, 1977; Lyons, 1977) or accessibility (e.g., Jarbou, 2010; Piwek et al., 2008). Here

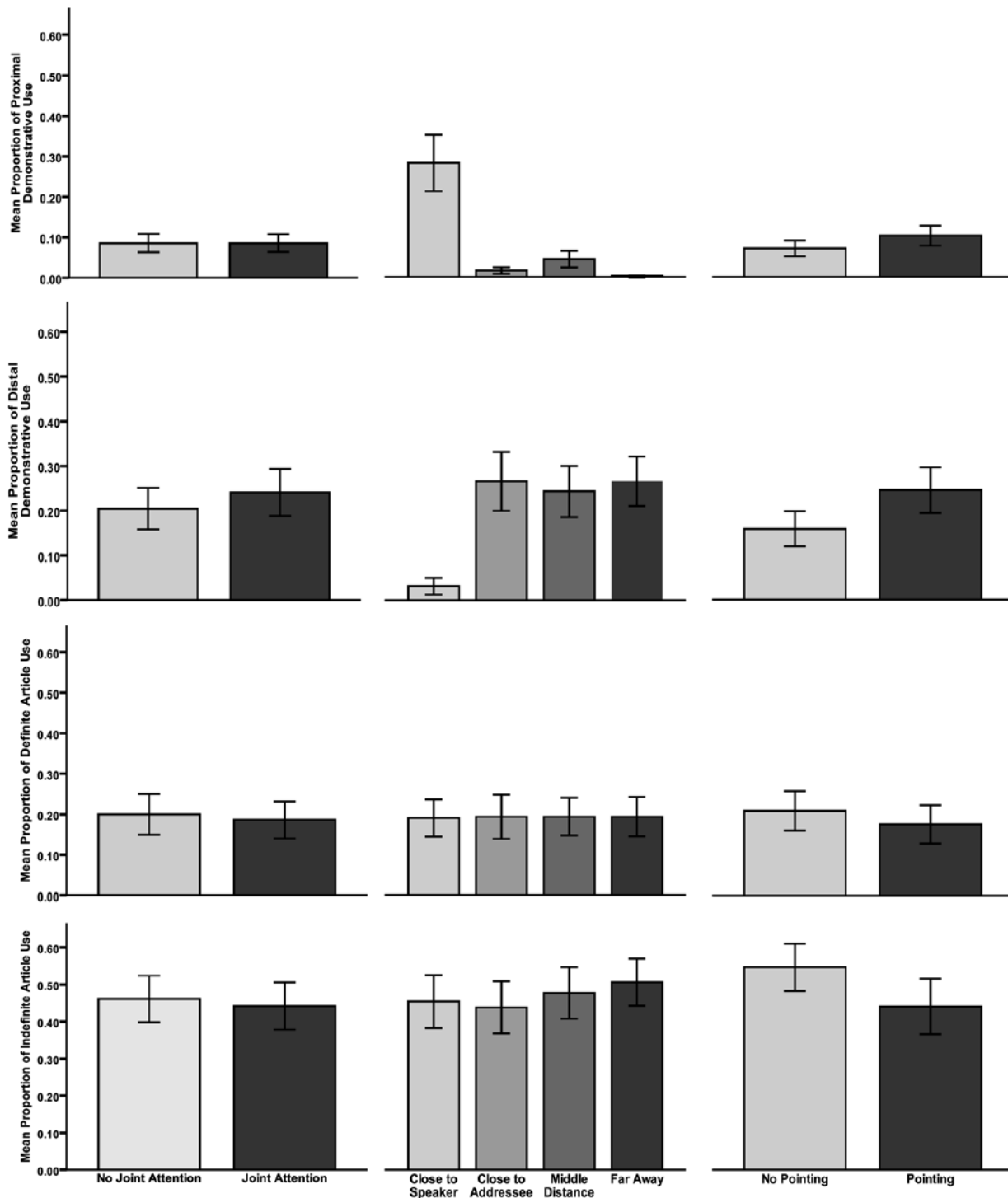


Figure 2: Separate panels for the mean proportion of use of proximal demonstratives, distal demonstratives, definite articles, and indefinite articles as a function of the three factors manipulated in the target pictures (Joint Attention, Location, and Pointing Gesture). Error bars represent standard errors of the mean.

we show that different factors may play a role and that speakers have different strategies at their disposal. Moreover, this finding underlines the importance of contrasting different factors in the same study (cf. Coventry et al., 2008; Stevens & Zhang, 2013). The current study further specifies that the influence of these factors may play a significant role in demonstrative choice, but not in the choice of (in)definite articles.

Finally, we found that both proximal and distal demonstratives were used more often with than without a pointing gesture. In contrast, indefinite articles were used more in the absence of a gesture. Bangerter (2004, p. 418) suggested that demonstratives may direct gaze to a concurrently used pointing gesture when the gesture carries the main informational burden. Our results are in line with this suggestion and underline that a pointing gesture may demarcate the addressee's search space, in our case when a distal demonstrative was used in reference to an object not near the speaker. Previous research has shown that speakers design the exact kinematic properties of their index-finger pointing gesture, such as its velocity and the duration of its post-stroke hold-phase, by taking into account the mental state of their addressee (Peeters et al., 2013). A demonstrative could then indeed be used to make the addressee pay attention to such an effort.

To conclude, the current study showed differential roles of visual joint attention, physical proximity, and the presence of pointing gestures in demonstrative choice. Our findings underline that the very basic human communicative ability of establishing triadic joint attention to a referent turns out to be dependent on a subtle interplay between different context-dependent factors, as reflected in one's choice of demonstrative.

References

- Anderson, S. R., & Keenan, E. L., (1985). Deixis. In T. Shopen (Ed.), *Language typology and syntactic description*. Cambridge: Cambridge University Press.
- Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, 24, 65-87.
- Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55, 1278-1289.
- Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, 15(6), 415-419.
- Burenhult, N. (2003). Attention, accessibility, and the addressee: The case of the Jahai demonstrative ton. *Pragmatics*, 13, 363-379.
- Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, 255, Vol. 63, 1-174.
- Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In I. A. Noveck & D. Sperber (Eds.), *Experimental Pragmatics*. Basingstoke: P. Macmillan.
- Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near-far perceptual space and spatial demonstratives. *Cognition*, 108, 889-895.
- Diessel, H. (1999). *Demonstratives. Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.
- Diessel, H. (2005). Distance contrasts in demonstratives. In M. Haspelmat, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The World Atlas of Language Structures*. Oxford: Oxford University Press.
- Enfield, N. J. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 82-117.
- Halliday, M. A. K., & Hasan, R. (1977). *Cohesion in English*. London: Longman Group Ltd.
- Hanks, W. F. (1990). *Referential practice: Language and lived space among the Maya*. Chicago: University of Chicago Press.
- Jarbou, S. O. (2010). Accessibility vs. physical proximity: An analysis of exophoric demonstrative practice in Spoken Jordanian Arabic. *Journal of Pragmatics*, 42(11), 3078-3097.
- Kirsner, R. S., & Van Heuven, V. J. (1988). The significance of demonstrative position in modern Dutch. *Lingua*, 76(2), 209-248.
- Küntay, A., & Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *Journal of Child Language*, 33, 303-320.
- Lyons, J. (1977). *Semantics. Volume 2*. Cambridge: Cambridge University Press.
- Peeters, D., Chu, M., Holler, J., Özyürek, A., & Hagoort, P. (2013). Getting to the point: The influence of communicative intent on the kinematics of pointing gestures. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1127-1132). Austin, TX : Cognitive Science Society.
- Piwek, P., Beun, R. J., & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, 40, 694-718.
- Senft, G. (2004). *Deixis and Demonstratives in Oceanic Languages*. Canberra: Pacific Linguistics.
- Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics*, 26, 31-45.
- Strauss, S. (2002). This, that, and it in spoken American English: a demonstrative system of gradient focus. *Language Sciences*, 24(2), 131-152.
- Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.
- Van Deemter, K., Gatt, A., van Gompel, R. P. G., & Kraemer, E. (2012). Toward a computational psycholinguistics of reference production. *Topics in cognitive science*, 4(2), 166-183.