Psychological proximity guides multimodal communication

Madeleine Long, Aslı Özyürek, Paula Rubio-Fernández

Max Planck Institute for Psycholinguistics

Madeleine.Long@mpi.nl, Asli.Ozyurek@mpi.nl, Paula.RubioFernandez@mpi.nl

Human communication is multimodal: from the use of speech to bodily articulators, language is expressed through multiple channels. However, only recently have researchers moved away from unimodal accounts of language to embrace its flexible and multimodal nature. Here we add to this growing body of work by examining two universal and early acquired means of reference: demonstratives and pointing gestures. While traditional accounts posit that demonstratives signal **spatial distance** [1], it has more recently been argued that they signal **psychological proximity** (i.e., whether the referent is in the shared space construed by interlocutors during conversation) [2]. In this view, speaker and listener jointly establish which referents are psychologically proximal, as indicated by many interactive factors (e.g., visibility, familiarity or ownership). Here we tested the roles of spatial distance and psychological proximity in the use of demonstratives and pointing by Turkish speakers.

Prior naturalistic observations in Turkish indicate that "şu" (the middle demonstrative form) is used to redirect the listener's attention towards the intended referent, accompanied by pointing [3]. This suggests that "şu" may be particularly sensitive to psychological proximity (in contrast to the proximal and distal demonstratives "bu" and "o"). Supporting this analysis, recent empirical work using still images revealed that "şu" is used more often when a speaker and listener's visual perspectives are misaligned [4]. Here we tried to replicate and extend this through more naturalistic materials (i.e., controlled videos and free interaction) to examine the roles of perspective alignment, pointing gestures and referent distance on demonstrative choice.

Experiment 1: 58 native Turkish speakers were shown 48 videos in which a speaker and a listener appeared on opposite sides of a table with four identical objects (for a sample display with a video link, see Fig. 1). Participants were asked to play the role of the speaker and request a target object (which appeared inside a circle) from the listener. To do so, participants completed the phrase in the speech bubble ("Now I need…") with either "bu", "şu" or "o" (equivalent to 'this one', 'that one' and 'that one over there' in English). We manipulated the position of the listener and the target (Positions 1-4), the listener's perspective (aligned or misaligned) and the presence or absence of a pointing gesture for a fully crossed design.

Using logistic mixed effects regression, we modelled Şu Choice (Şu=1, Bu/O=0) with Pointing (Pointing vs No Pointing), Position (1-4) and Visual Perspective (Aligned vs Misaligned) as fixed effects with maximal RE structure. As predicted, there were main effects of Pointing (β=.911, SE=.177, p<.001), Visual Perspective (β=2.100, SE=.272, p<.001) and Position (β=.408, SE=.069, p<.001) with şu being selected more often with pointing, in misaligned trials and in the middle positions (2-3). There were also Pointing x Position (β=.255, SE=.115, p=.027) and Position x Visual Perspective interactions (β=.517, SE=.158, p=.001) whereby Pointing and Visual Perspective had greater influence on şu choice in more distant positions (Figs. 2 and 3), perhaps because in close proximity attention correction is less marked (including the use of pointing) and interlocutors rely on subtle eye/head movements instead. Importantly, LMER models for Bu and O did not reveal greater choice in misaligned trials, supporting the notion that şu is used for attention correction in Turkish.

Experiment 2: To increase ecological validity, and further explore multimodal behavior when perspectives are misaligned, we used a free form multimodal interactive task where the participant wore eye-tracking glasses and a confederate acted as the listener (N=5; data collection ongoing). In half the trials, the speaker *directed* the listener's attention to the target and in the other half the listener guessed the wrong target (by looking two objects away), leading the speaker to *correct* the listener (Fig. 4).

Replicating the video results, there was a preference for şu and pointing in *correction* trials. Şu was also linked to the use of negation (for correction) and more frequent fixations to the listener's face compared to bu and o (Table 1), likely reflecting a monitoring of the listener's perspective to redirect attention to the target object. Interestingly, distinct patterns emerged in the type of pointing gesture used (i.e., pointer finger, middle finger, or full hand) depending on whether the speaker *directed* or *corrected* attention and what demonstrative they produced (Fig. 5). Şu was associated with the greatest variety of pointing types, especially in attention *correction* trials, where all three types were used. One explanation for this variation is that in addition to foregrounding information (i.e., locating the target for the listener) gesture choice can signal secondary information, such as a lack of referential common ground [5]. This would explain why the middle finger was used exclusively with şu for attention *correction*, as the target was noticeably not in common ground. Thus, the middle finger may have served a dual-purpose, signalling the target location (i.e., Position 2 or 3), while also operating as an iconic gesture, indicating that the target was in between other similar objects. Another interesting pattern that emerged from our eye-tracking data was that participants used three distinct perspective-taking strategies during attention *correction* trials (Table 2). These strategies involved direct and indirect ways of assessing whether visual perspectives were aligned (e.g., by directly following the listener's gaze to the incorrect target (then correcting them), or by simply looking at their eyes and calculating that their guess was incorrect. As more data comes in, we will test how these strategies relate to pointing and demonstrative choice.

Together, our results support prior naturalistic [3] and empirical work [4] demonstrating that the Turkish middle demonstrative is used for attention correction. Turkish speakers thus dynamically integrate demonstrative use and pointing as a function of both **spatial distance** and **psychological proximity**. Building on recent cross-linguistic work [4], this lends support to the notion that different demonstrative systems pose different demands on social cognition, reflected both in the choice of demonstrative form and the type of pointing gesture used, which can convey additional pragmatic information.

Index Terms: pointing gestures, demonstratives, psychological proximity

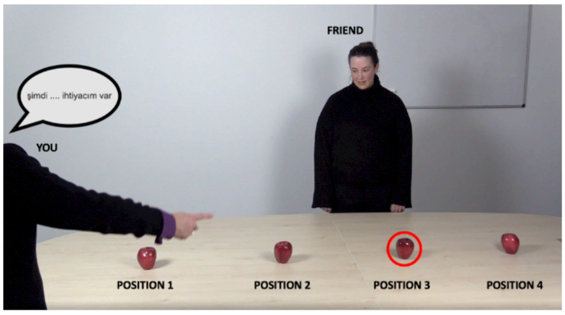

Figure 1: Exp 1 trial with misaligned perspectives and pointing. Link to full video: https://osf.io/s7dq3

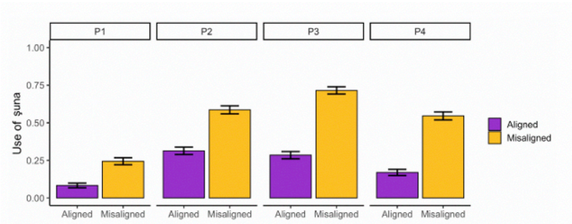

Figure 3: Exp 1 Position x Visual Perspective interaction.

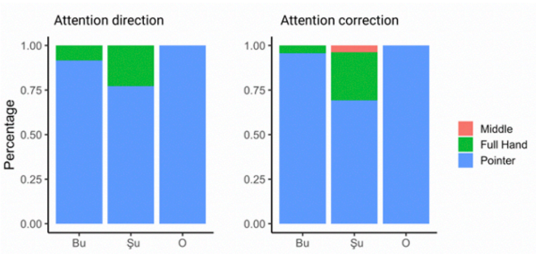

Figure 5: Exp 2 breakdown of pointing gesture type by trial type (attention direction or attention correction) and demonstrative form.

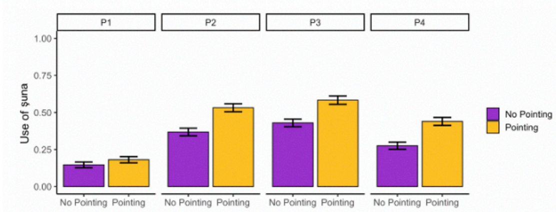

Figure 2: Exp 1 Pointing x Position interaction.

Figure 4: Still image from synched cameras in Exp 2. The top images show the video camera perspectives, the bottom is the perspective from the eye-tracking glasses.

Demonstrative	Looks to the face	Negation
Bu	M= 0.957	M= 0.277
Şu	M= 1.128	M= 0.417
O	M= 1	M= 0.333

Table 1: Descriptive statistics for Exp 2 pilot data. Note: 'Looks to the face' refers to the number of looks to the face leading up to and at the point of demonstrative use.

Perspective-taking strategies	
Strategy 1	Look at the listener's face → check the computer screen for the target object → note that they are looking at the wrong object and redirect their attention to the target.
Strategy 2	Look at listener's face and immediately note that they are looking at the wrong object → redirect their attention to the target.
Strategy 3	Look at listener's face → follow their eye gaze to the wrong object → note that they are looking at the wrong object and redirect their attention to the target.

Table 2: Perspective-taking strategies based on eye-tracking data from Exp 2 attention correction trials.

References

[1] Diessel, H., & Coventry, K. R. (2020). Demonstratives in spatial language and social interaction: An interdisciplinary review. *Frontiers in Psychology*, *11*, 555265.

[2] Peeters, D., & Özyürek, A. (2016). This and that revisited: A social and multimodal approach to spatial demonstratives. *Frontiers in Psychology*, *7*, 222.

[3] Özyürek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-face conversational interaction. In S. Santi, I. Guaitella, C. Cave, & G. Konopczynski (Eds.), *Oralite et gestualite: Communication multimodale, interaction: actes du colloque ORAGE 98* (pp. 609-614). Paris: L'Harmattan.

[4] Rubio-Fernández, P. (2023). Demonstrative systems: From linguistic typology to social cognition. *Cognitive Psychology, 139,* 101519.

[5] Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics, 39*(10), 1722-1741.