

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Anticipating Naming: Multimodal Behavioural Dynamics Leading to Successfully Attended Naming Events

Permalink

<https://escholarship.org/uc/item/3hm1x2bv>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Sander, Jennifer
Eikelboom, Dilys
Zhang, Yayun
[et al.](#)

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Anticipating Naming: Multimodal Behavioural Dynamics Leading to Successfully Attended Naming Events

Jennifer Sander (jennifer.sander@mpi.nl)

Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands
Max Planck School of Cognition, Stephanstraße 1A, 04103 Leipzig, Germany

Dilys Eikelboom

Maastricht University, Oxfordlaan 55, 6229EV Maastricht, The Netherlands

Yayun Zhang

Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands

Caroline F. Rowland

Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands
Donders Institute for Brain, Cognition and Behaviour, Radboud University, Thomas van Aquinostraat 4, 6525 GD, Nijmegen, The Netherlands

Abstract

What leads to successfully attended naming events? To map novel labels onto novel objects, children must attend to the correct objects around during naming events. The current study investigates, through a bottom-up approach, which behavioural dynamics lead to successfully attended naming events. In a semi-naturalistic head-mounted eye-tracking study, we tested 23 caregiver-child dyads (17 signing, 19 non-signing). Children acquiring a sign language are known to distribute their visual attention differently from children acquiring a spoken language. The current study investigates if differences in language modality lead to differences in the behavioural dynamics of attended and unattended naming events. Findings suggest that successfully attended naming events follow different behavioural dynamics than unattended naming events and that effects of language modality are most prominent in the amount of mutual gaze between children and their caregivers.

Keywords: Sign Language; NGT; Word Learning; Concept Acquisition; Visual Attention; Head-Mounted Eye-Tracking

Introduction

Naming events are of unique importance for children on their way to acquire language: they offer the opportunity to build label-object mappings; i.e., to connect a linguistic form to a meaning. Previous research has shown that the frequency with which caregivers name objects influences children's concurrent and subsequent vocabulary acquisition (e.g., Goodman, Dale & Li, 2008) but evidence suggests that not only the raw frequency of the language input itself, but also the frequency of successfully attended language input correlates with children's ability to map labels onto referents (Cartmill et al., 2013; Schroer & Yu, 2023). However, it is not yet clear what leads to successfully attended naming events. In the current study, we explore the behavioural dynamics leading to successfully attended naming events in caregiver-child dyads using a spoken or sign language.

A successfully attended naming event could, for example, be an instance during which a caregiver names the object the child is currently playing with and attending to via gaze and touch, allowing for the mapping between object and label to occur (see studies on simultaneous label-object input, e.g.,

Bothe et al., 2024; Gogate & Bahrick, 1998; Gogate, Bahrick & Watson, 2000; Yu & Smith, 2012; Perry et al., 2022). Understanding the behavioural patterns associated with an increased likelihood that children pay attention to a naming event is crucial for understanding the facilitatory effect of language input on children's word learning.

Naturalistic caregiver-child interactions are highly complex, high dimensional, multimodal events, which makes it difficult to track and explore the dynamics of the behavioral patterns that lead to successful label-object mappings. That said, a number of different attentional behaviours, carried out by the child or their caregiver, have been found to be associated with successfully attended naming events either in isolation or in combination:

- **Visual attention** in the form of gaze to the object, following the gaze of the interlocutor, gaze overlap and mutual gaze between interlocutors (e.g., Yu & Smith, 2012; Yu et al., 2019)
- **Tactile attention** in the form of holding, touching or manipulating an object (e.g., Booth et al., 2008; Yu & Smith, 2012 & 2013; Schroer & Yu, 2023)
- **Attention behaviours** and deictic gestures (e.g., waving, pointing; tapping; e.g., Deák et al., 2018)
- **Referential language use** (e.g. demonstratives in spoken and pointing in sign language; Leary et al., 2022 for ASL; Chang & Deák, 2019 for English)

Each of these behaviours has been successfully associated with children's word learning abilities. For example, Schroer and Yu (2023) showed that children who attended to named objects with gaze and touch more frequently in the assessed play session had better learning outcomes of the novel labels.

While these behaviours have been shown to be important, they may look different in interactions using a sign language, since language modality affects the level of engagement with these behaviours. For example, in spoken interactions, children rarely gaze at their caregiver but are more likely to follow their caregiver's manual actions (Yu & Smith, 2013). In contrast, children in sign language interactions have been

observed to frequently engage in mutual gaze with their caregivers, likely because these interactions require visual attention for the sign to be successfully perceived (Gale & Schick, 2009; Lieberman et al., 2014). Hence, differences in the behavioural dynamics of signed and spoken interactions are likely and need to be considered when investigating what leads to successfully attended naming events.

To explore these behavioural dynamics in a way that can account for differences between interactions using spoken and sign languages, we used a bottom-up data-driven approach. Bottom-up approaches invert the usual research process by putting the raw data first, allow the systematic exploration of patterns in the unfiltered behavioural data, potentially resulting in the emergence of new hypotheses (as opposed to defining a psychological construction (e.g., joint attention) and operationalising it using pre-defined criteria, e.g., defining joint attention as gaze overlap or coordinated attention, see Sander et al., 2024). For example, Pereira and colleagues (2014) conducted a bottom-up approach on visual attention behaviours, and found that visual dominance of a named target object was associated with learning. Similarly, a study on gaze and object touch by Sun and Yoshida (2022) found that children were more likely to sustain visual attention to the object if their caregiver looked at the same item and produced a naming event, compared to when the caregiver used other linguistic cues (e.g. calling the child's name). However, no bottom-up studies we are aware of account for potential differences in language modality.

The aim of the current study was to explore the following question: which patterns of behaviour are associated with an increased likelihood that caregiver naming events are attended to by the child, and thus become effective opportunities for the child to build successful label-object associations in speaking and signing dyadic interactions? We took advantage of recent technological advances, enabling us to collect head-mounted eye-tracking data from young children, to explore the data from a first-person perspective. This allowed us to identify what is in the visual field of child and caregiver during an object labelling event.

We investigated which behaviours, and combinations of behaviours, lead to successfully attended naming events, (i.e. those events that are likely to enable children to make successful label/object mappings), in a three-step system: First, we identified in which way, and by use of which behaviours, children attended to their caregivers or to the named object during successfully attended naming events. We wanted to understand if children attended to a naming event by engaging with the object or caregiver, through gaze, touch, other attentional behaviours, or a combination of several behaviours. Second, we identified the most frequent patterns of behaviours in the three seconds leading up to the naming event, to account for findings reporting that changes in behaviour related to the occurrence of a naming event start around 3 seconds before a naming event (Pereira et al., 2014). Finally, we assessed the temporal dynamics of visual and tactile behaviours in the pre-naming time window. In all three steps, we compared sign and spoken language modality.

Method

This study was part of a larger study conducted with the purpose of investigating how social interactions in signing and non-signing dyads may facilitate word learning. Find more information about the study here: <https://osf.io/ha57v/>

Participants

Participants were 23 children (11 girls) between one and five years of age (mean age = 40.87 months, SD = 14.62 months, range = 13-63 months) and their primary caregivers. Dyads were either interacting in Sign Language of the Netherlands (NGT, N = 10 signing dyads + 2 Bimodal Bilinguals) or in spoken Dutch (NL, N = 11 speaking dyads + 2 Bimodal Bilinguals) throughout their sessions. A subset of dyads (N = 13 dyads), participated at two timepoints, about six months apart (mean = 6.09 months, range = 5.55 – 6.68 months). For two children (bimodal bilinguals), data was collected in both language modalities. The complete dataset consisted of 42 toy play sessions across 23 children.

Materials and Procedures

Signing dyads were recruited from across the whole of the Netherlands and tested in a mobile lab, to allow for maximal flexibility in testing location. Dyads using a spoken language were recruited through the recruitment database of the local babylab and were tested in the same mobile lab.

Children played with their caregiver in a naturalistic toy play situation, while both interaction partners were wearing a head-mounted eye tracker with self-calibrating software (Pupil Labs NEON). The scene was additionally filmed using two fixed GoPro (Hero 9). Toys provided included a cash register, a number of wooden and plastic food items, a few plushies, balls, wooden trees, and a number of stackable boxes. Children and caregivers were seated at a 90-degree angle to each other at a table, or children were seated on their caregiver's lap if they were not comfortable with the other seating arrangement (N = 5). Caregivers were asked to play with their children and the toys, and use language as they normally would. Each play session lasted around 10 minutes.

Processing

All four video streams (eye-tracking for each participant and two fixed camera perspectives) were synchronised using video-editing software. Data was then annotated in the audio-visual annotation program ELAN (Version 6.4; MPI for Psycholinguistics, The Language Archive, 2022). All signed and spoken interactions were annotated. NGT was annotated by two fluent signers (one deaf, one hearing signer), spoken Dutch was first annotated using Whisper (OpenAI, 2022) and then corrected by fluent Dutch speakers.

Based on the language annotations, we identified all **caregiver naming events** that labelled an object present in the scene: these were defined as events that contained a label (noun) produced by the caregiver that referred to one of the objects present, in either spoken or sign language (e.g. *Money*, but also *Cash*, or *Coins*). The onset of a signed naming event

was defined as the onset of the first manual movement that was part of the sign (e.g. hand raise) and the offset of the naming event was defined as the offset of the last manual movement that was part of the sign (e.g. lowering hand). For spoken naming events, the onset was defined as the first phone of the noun and the offset was then defined as the offset of the last phone of the noun. Naming events were considered successfully attended to if the child directed any behaviour (gaze, touch, attention behaviour or referential language) at the object or at their caregiver during the naming event. As signers have to attend to their interaction partner to perceive a sign, we included looks from the child to the co-participant as well as any behaviour directed at the object into our considerations for successfully attended naming events. In addition to the caregivers' naming events, we annotated the following behaviours centered on the named target.:

- **Visual attention** measured by gaze location (based on head-mounted eye-tracking), differentiating between gaze to named objects, gaze to the interaction partner, mutual gaze and gaze overlap on objects
- **Tactile attention** (= touch) to the named object
- Object-directed **attentional behaviours** (e.g., tapping object, reaching for object) throughout the whole session for both interaction partners
- **Referential language use** by both interaction partners (for signers, we considered all pointing events as referential language use, for non-signers we considered all use of referential language in the form of demonstratives, e.g. *dit*, *deze*, [this] and *dat*, *die* [that])

Based on the identified naming events, we automatically extracted a fixed pre-naming window of 3 seconds and assessed the patterns of behaviour during this time window.

Results

All analyses were conducted in R (R Core Team, 2023), with the significance level set to .05 (two-tailed), though, we also indicate if $p < .01$ and $p < .001$. The assumptions of all statistical tests were met without transformations. The data, processing and analysis code as well as supplementary materials can be found here: <https://osf.io/7sxc4/>

How children attend to naming events

After identifying successfully attended naming events, we assessed the frequency of each behaviour with which children demonstrated their attention to naming events, and whether the behaviours with which the child attended to the naming event differed by language modality. To do so, we conducted separate generalized linear mixed effects models, assessing the effect of language modality on the presence of each behaviour. The models also included random effects for participants, and session (if this did not lead to singularity issues: behaviour presence $\sim$ Language Modality + (1|Participant) + (1|Session)), Results are visualised in Figure 1. Overall, the behaviours involved were similar across the two language modalities: gaze was involved in the majority of naming events, followed by object touch (effect of language modality was n.s.). Only a small number of naming

events across both modalities were accompanied by children's referential language or object-directed attention behaviours (effect of language modality again n.s.). The most common behaviour combination with which children attended to naming events across both language modalities was touch combined with gaze to object. However, children using spoken Dutch attended to naming events significantly more often by gaze to the object ($\beta = 0.61$, $SE = 0.18$, $z = 3.56$, $p < .001$), and by gaze overlap onto the object with their caregiver ($\beta = 0.64$, $SE = 0.23$, $z = 2.81$, $p < .01$) than children using NGT. There were also large differences in the number of gazes to the caregiver between the two groups of children: children using sign language attended significantly more often to naming events by gazing at their caregiver ($\beta = -2.78$, $SE = 0.29$, $z = -7.98$, $p < .001$) and by mutual gazing at each other than children using spoken language ($\beta = -2.18$, $SE = 0.3$, $z = -7.35$, $p < .001$).

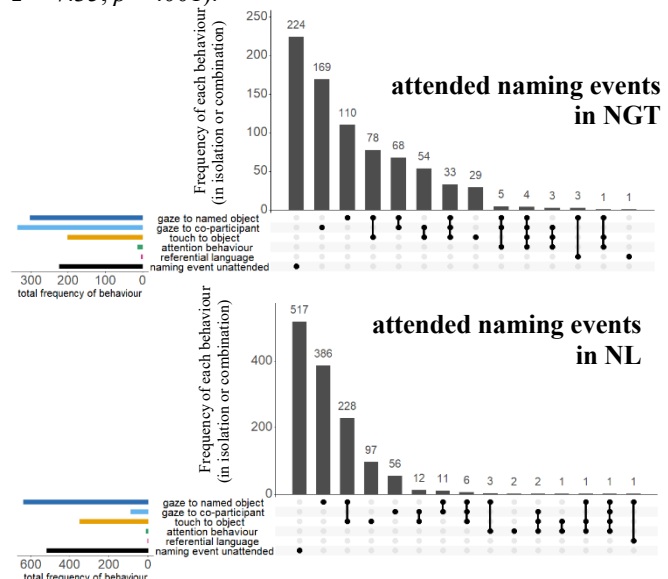


Figure 1: Visualisation of the frequency of the behaviour (combinations) with which children successfully attend naming event, split by language (NGT: top, NL: bottom). Coloured bars (left) show the total frequency of each behaviour (both in isolation and in combination), black bars (right) show frequency of individual behaviours and behaviour combinations, as indicated by the dots in the bottom right.

What leads to naming events

After establishing how children attend to naming events, we assessed which behaviours and behaviour combinations by both caregiver and child lead up to naming events, across language modalities.

We assessed this separately for successfully attended and unattended naming events (those in which the child did not direct a behaviour to object or caregiver), split by language. We assessed, for each naming event in a three second pre-naming time window, whether a specific behaviour or behaviour combination was present. We observed that the most frequent behaviour constellations leading to attended naming events were multi-behaviour patterns (e.g. *caregiver gaze to object + caregiver gaze to child + child gaze to object*

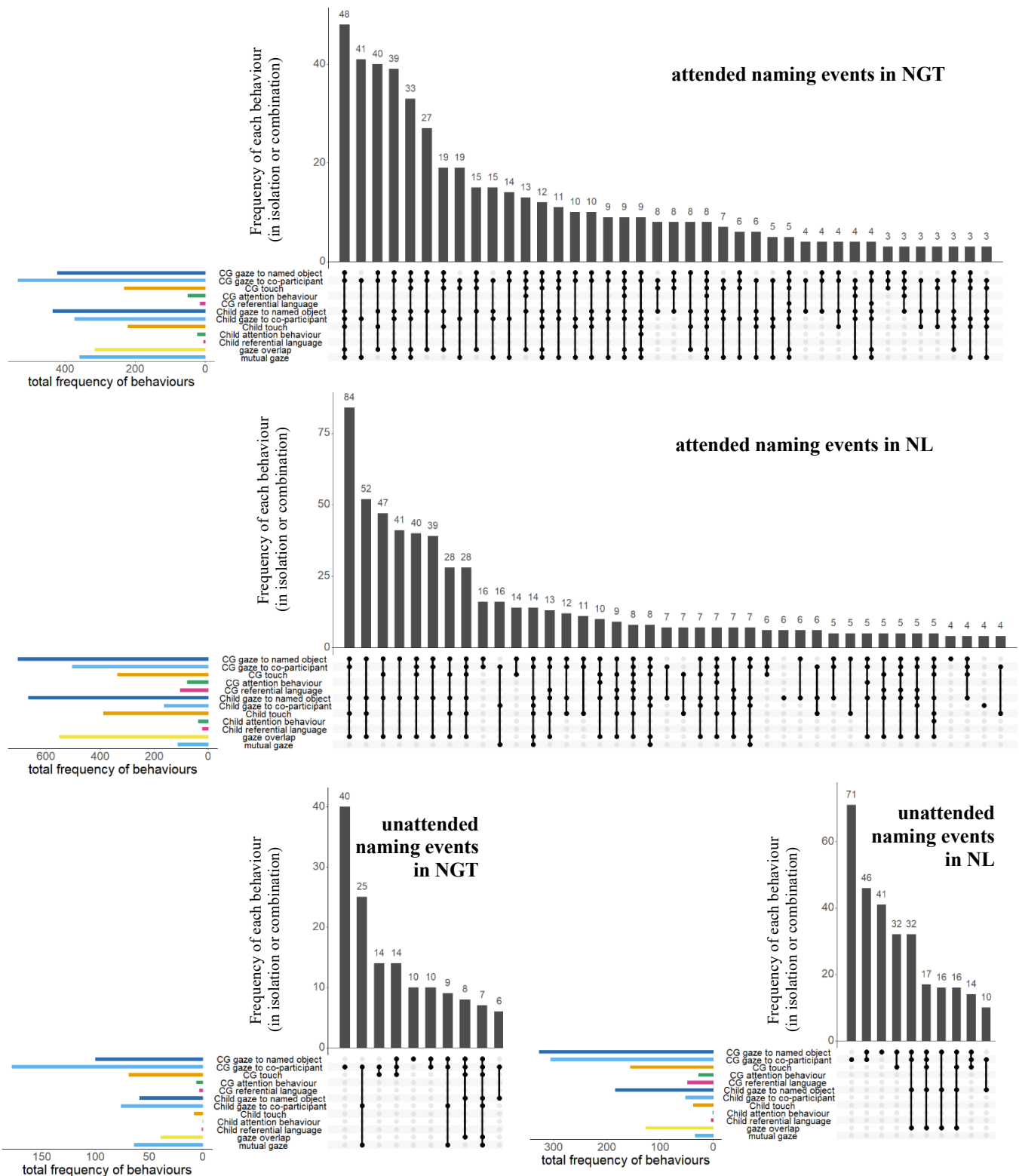


Figure 2: Visualisation of frequency of behaviours that children and caregivers (CG) engage in during the pre-naming time window of successfully attended naming events by language (top two visualisations, NGT: top, NL: bottom) and top ten unattended naming events by language (bottom row visualisations, NGT: left, NL: right). For reasons of space, and since we are primarily interested in identifying behaviours that lead to attended naming events we show all behaviours for attended, but only the top 10 for unattended events. Coloured bars (left in each visualisation) show the frequency of each behaviour individually, black bars (right within each visualisation) shows frequency of behaviours and behaviour combinations, as indicated by the dots in the bottom right.

+ *child gaze to caregiver* + *mutual gaze* was the most common combination leading to attended sign naming events, see Figure 2). In contrast, the most frequent behaviours and behaviour combinations leading to unattended naming events were mono-behaviours, behaviour combinations consisting of only a few behaviours, and often only involving behaviours by the caregiver (see Figure 2). This was true for both naming events in NGT and in NL. Differences between languages were again observed in the frequency with which mutual gaze was part of the pre-naming time window for attended naming events; signers gazed more at each other and spent more time in mutual gaze than non-signers.

We assessed these findings using a generalized linear mixed effects model, predicting whether a naming event was attended to by the number of relevant behaviours observed in the pre-naming time window and by whether mutual gaze was present in the pre-naming time window, including an interaction with language modality. The models also included a random effect for participants (naming event attendance ~ number of relevant behaviours observed in pre-naming time window + mutual gaze*Language Modality + (1|Participant)). We found a significant effect for the number of behaviours ($\beta = 0.73$, $SE = 0.04$, $z = 18.75$, $p < .001$), and a significant interaction between the presence of mutual gaze and language modality ($\beta = -0.67$, $SE = 0.34$, $z = -2.06$, $p < .05$) but no overall effect of language modality or mutual gaze. The interaction suggests that mutual gaze was only a relevant predictor for signers but not for non-signers.

Temporal dynamics of behaviours in pre-naming time window

To better understand what leads to successfully attended naming events, we investigated the temporal dynamics of all child and caregiver behaviours of visual and tactile attention, as these behaviours can be measured in terms of duration, during the pre-naming time window of three seconds before naming onset. We differentiated between successfully attended and unattended naming events and whether the behaviour was directed at the named object or the co-participant, focusing on behaviours of the child or both interaction partners. For each assessed behaviour, we measured how often it occurred at each frame (77 ms time bin) within the 3000 ms pre-naming time window leading up to a naming event, expressed as the proportion of all naming events (successful vs. unsuccessful, signed vs. spoken) in which it appeared. Additionally, we visualize the 1000 ms after naming onset, which covers the majority of naming event durations in our study (mean naming event duration: 506ms, $SD = 348$ ms). Results are illustrated in Figure 3.

We assessed the temporal envelope by employing cluster-based permutation analysis on the comparison of each behaviour between attended and unattended events, and for gaze, between gaze to objects and co-participants.

We found significant effects for child gaze across both languages: children looked at the named object significantly more often in the context of attended compared to unattended naming events throughout both, pre-naming period and

naming period. They looked significantly more to their caregivers shortly before naming onset and during naming events that were attended compared to unattended naming events. They also looked more to caregiver compared to objects right before naming onset in unattended naming contexts (NL only), and more to the caregiver compared to objects in the pre-naming period (NGT) and in pre-naming and naming period of attended objects (NL). In regards to mutual gaze, children in both language modalities spent more time in mutual gaze with their caregivers shortly before onset and during attended compared to unattended naming events. They also shared significantly more gaze to the named objects with their caregivers throughout the whole pre-naming and naming period of attended compared to unattended naming events, in both language modalities. Finally, throughout the whole pre-naming and naming period children in both language modalities spent significantly more time touching the named object in attended compared to unattended naming events.

Differences between both languages were striking, especially in the amount of looks to the caregiver: Signing children looked equally often at objects and their caregiver during the 1000 ms before and after naming event onset, while non-signing children rarely looked at their caregivers at all.

Discussion

In the current study we assessed which patterns of behaviour increased the likelihood that caregiver naming events were attended by a child, allowing the child to build successful label-object associations. We investigated this in three sets of analyses. We found that successfully attended naming events were more often preceded by multi-behavioural (vs. mono-behavioural) pre-naming patterns, and in signers, by higher levels of mutual gaze. Successfully attended events differed from unattended events in the temporal envelope of the pre-naming period across most measures of visual and tactile attention, with clear peaks about one second before naming onset. Language modality further shaped these patterns, partly leading to reversed preferences (e.g. preferences to look at co-participants over objects in signers vs. preferences to look at objects over the co-participant in non-signers).

Overall, the results of the current study suggest that interactional behavioural patterns start to vary already in the pre-naming period, before the actual naming event onset, and while these patterns have some general characteristics that apply across modalities (e.g. multi-behavioural pre-naming periods precede successfully attended naming events more frequently than mono-behavioural periods), we also observed language modality-specific patterns (e.g. in the frequency and amount of mutual gaze involvement in pre-naming and naming periods between signing and non-signing dyads).

By focusing on behaviours preceding naming events, the study captures interactional patterns linked to the caregivers' decision to name and children's successful attendance of these naming events. We acknowledge, however, that the pre-naming window represents only a subset of the broader

temporal dynamics of naming interactions, particularly those in which caregivers follow in on objects the child is already attending to. Other types of naming events may reflect parents' attempts to re-direct children's attention, with relevant cues emerging primarily in the post-naming period. These events can also be successful learning moments if the caregiver's naming or subsequent behaviours draws the child's attention to the referent. Therefore, successful naming can arise through multiple pathways, either by reacting to children's current focus or by actively guiding their attention. Future research is necessary to disentangle the temporal

dynamics of naming events, their pre- and post-naming periods as well as the dynamic between attention to naming events and the choice of a caregiver to name.

The findings of the current study emphasize the richness of naturalistic interactions and illustrates the advantage of data driven approaches. The behavioural patterns and temporal dynamics identified in the current study offer the opportunity for future studies to generate new research questions and to investigate the relationship between these behavioural patterns and temporal dynamics and children's word learning abilities more closely.

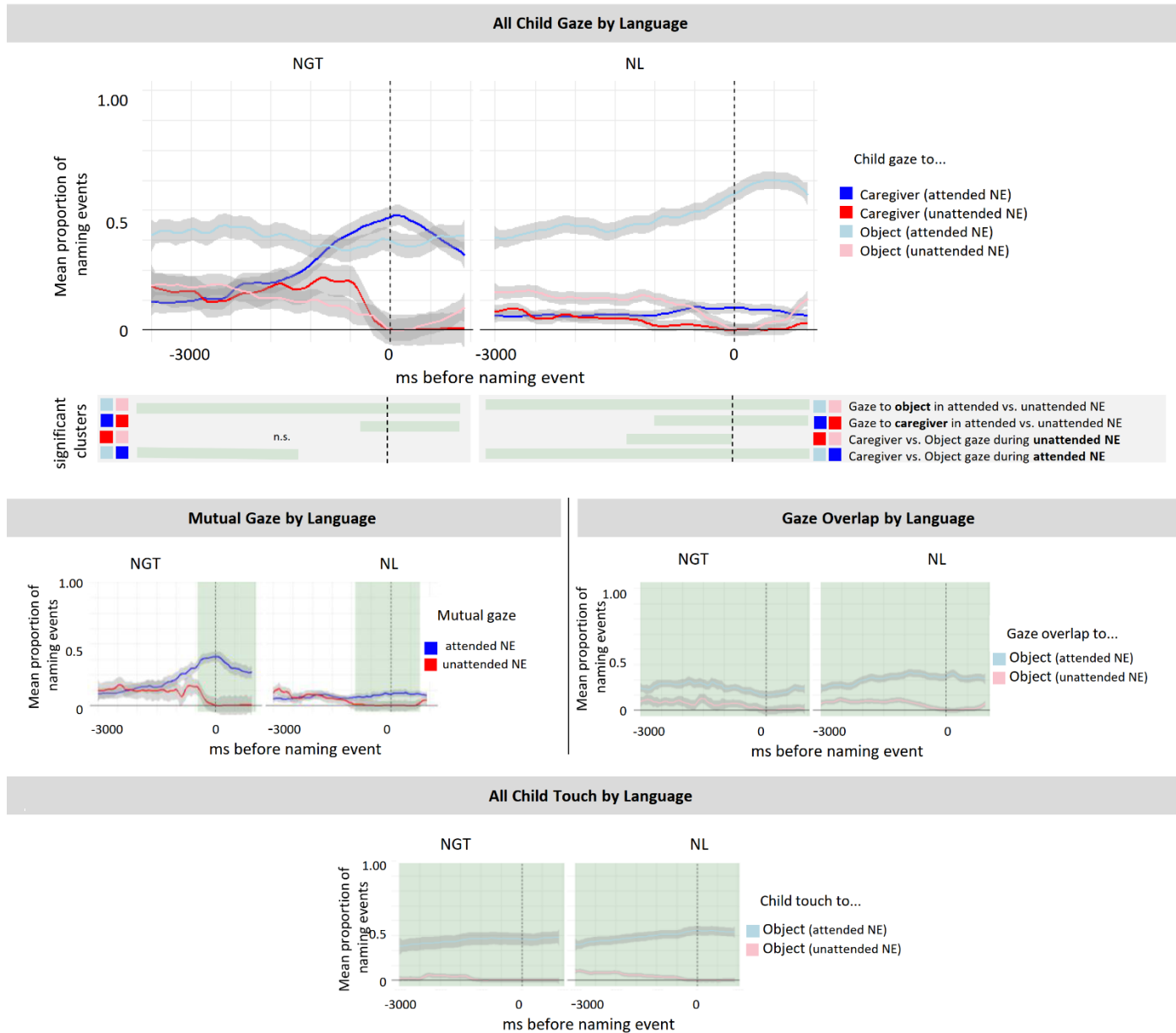


Figure 3: Temporal dynamics of proportion of naming events with behaviours present (y axis: proportion from 0 to 1) in a three second pre-naming time window (x-axis in ms, from -3000 ms before to 1000 ms after onset of naming), by Language (NGT: left, NL: right in each plot). Behaviours: Child gaze (top), mutual gaze (middle, left), gaze overlap (middle, right), Child touch (bottom). Significant differences between lines at different timepoints as identified by cluster-based permutation analysis indicated by green shading. For gaze, several comparisons were drawn and significant time windows are indicated by the green stripes below the gaze plots.

References

- Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-Pragmatics and Attention: Contributions to Gesturally Guided Word Learning in Toddlers. *Language Learning and Development*, 4(3), 179–202. <https://doi.org/10.1080/15475440802143091>
- Bothe, R., Eiteljoerge, S., Trouillet, L., Elsner, B., & Mani, N. (2024). Better in sync: Temporal dynamics explain multisensory word-action-object learning in early development. *Infancy*, 29(4), 482–509. <https://doi.org/10.1111/infa.12596>
- Cartmill, E. A., Armstrong, B. F., III, Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, 110(28), 11278–11283. <https://doi.org/10.1073/pnas.1309518110>
- Chang, L. M., & Deák, G. O. (2019). Maternal discourse continuity and infants' actions organize 12-month-olds' language exposure during object play. *Developmental science*, 22(3), e12770. <https://doi.org/10.1111/desc.12770>
- Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy*, 23(1), 4–28. <https://doi.org/10.1111/infa.12204>
- Gale, E., & Schick, B. (2009). Symbol-infused joint attention and language use in mothers with deaf and hearing toddlers. *American Annals of the Deaf*, 153(5), 484–503. <https://dx.doi.org/10.1353/aad.0.0066>
- Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, 69(2), 133–149. <https://doi.org/10.1006/jecp.1998.2438>
- Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, 71(4), 878–894. <https://doi.org/10.1111/1467-8624.00197>
- Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, 35(3), 515–531. <https://doi.org/10.1017/S0305000907008641>
- Leary, J., Gappmayr, P., & Lieberman, A. M. (2022). Pointing in parent input during interactions with deaf children in American Sign Language. In *Proceedings of the 46th annual Boston University Conference on Language Development* Cascadilla Press.
- Lieberman, A. M., Hatrak, M., & Mayberry, R. I. (2014). Learning to look for language: Development of joint attention in young deaf children. *Language Learning and Development*, 10(1), 19–35. <https://doi.org/10.1080/15475441.2012.760381>
- Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic bulletin & review*, 21(1), 178–185.
- Perry, L. K., Custode, S. A., Fasano, R. M., Gonzalez, B. M., & Valtierra, A. M. (2022). Coordination of caregiver naming and children's exploration of solid objects and nonsolid substances. *Frontiers in psychology*, 13, 945664. <https://doi.org/10.3389/fpsyg.2022.945664>
- Sander, J., Çetinçelik, M., Zhang, Y., Rowland, C. F., & Harmon, Z. (2024). Why does joint attention predict vocabulary acquisition? The answer depends on what coding scheme you use. In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Meeting of the Cognitive Science Society* (pp. 1607–1613). Cognitive Science Society. <https://escholarship.org/uc/item/3dm358j8>
- Schroer, S. E., & Yu, C. (2023). Looking is not enough: Multimodal attention supports the real-time learning of new words. *Developmental Science*, 26(2), e13290. <https://doi.org/10.1111/desc.13290>
- Sun, L., & Yoshida, H. (2022). Why the parent's gaze is so powerful in organizing the infant's gaze: The relationship between parental referential cues and infant object looking. *Infancy*, 27(4), 780–808. <https://doi.org/10.1111/infa.12475>
- Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244–262. <https://doi.org/10.1016/j.cognition.2012.06.016>
- Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS one*, 8(11), e79659. <https://doi.org/10.1371/journal.pone.0079659>
- Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, 22(1), Article e12735. <https://doi.org/10.1111/desc.12735>