UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Analysing Cross-Speaker Convergence in Face-to-Face Dialogue through the Lens of Automatically Detected Shared Linguistic Constructions

Permalink

https://escholarship.org/uc/item/43h970fc

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors

Ghaleb, Esam
Rasenberg, Marlou
Pouw, Wim
et al.

Publication Date

2024

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

Analysing Cross-Speaker Convergence in Face-to-Face Dialogue through the Lens of Automatically Detected Shared Linguistic Constructions

Esam Ghaleb (e.ghaleb@uva.nl)
Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands

Marlou Rasenberg (marlou.rasenberg@meertens.knaw.nl)
Meertens Institute, Royal Netherlands Academy of Arts and Sciences, The Netherlands

Wim Pouw (wim.pouw@ru.nl) & **Ivan Toni (ivan.toni@ru.nl)**
Donders Institute for Brain, Cognition, and Behaviour, Radboud University, The Netherlands

Judith Holler (judith.holler@mpi.nl) & **Aslı Özyürek (asli.ozyurek@mpi.nl)**
Donders Institute for Brain, Cognition, and Behaviour, Radboud University & MPI for Psycholinguistics, The Netherlands

Raquel Fernández (raquel.fernandez@uva.nl)
Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands

Abstract

Conversation requires a substantial amount of coordination between dialogue participants, from managing turn taking to negotiating mutual understanding. Part of this coordination effort surfaces as the reuse of linguistic behaviour across speakers, a process often referred to as *alignment*. While the presence of linguistic alignment is well documented in the literature, several questions remain open, including the extent to which patterns of reuse across speakers have an impact on the emergence of labelling conventions for novel referents. In this study, we put forward a methodology for automatically detecting *shared lemmatised constructions*—expressions with a common lexical core used by both speakers within a dialogue—and apply it to a referential communication corpus where participants aim to identify novel objects for which no established labels exist. Our analyses uncover the usage patterns of shared constructions in interaction and reveal that features such as their frequency and the amount of different constructions used for a referent are associated with the degree of object labelling convergence the participants exhibit after social interaction. More generally, the present study shows that automatically detected shared constructions offer a useful level of analysis to investigate the dynamics of reference negotiation in dialogue.

Keywords: face-to-face dialogue; referential communication; linguistic alignment

Introduction

Speakers use language flexibly in conversation in order to negotiate mutual understanding and achieve joint goals (Clark, 1996). Repeated referential games have provided a framework to study such coordination processes (Clark & Wilkes-Gibbs, 1986; Fay et al., 2018; Krauss & Weinheimer, 1966; Motamedi et al., 2019). These games typically involve several rounds where one participant (the director) describes a novel object to another one (the matcher), who needs to identify it among several candidates. In these referential director-matcher games, pairs are faced with two major challenges. First, the director and matcher need to work together so that the matcher can find the target object, which is a collaborative undertaking (Clark & Wilkes-Gibbs, 1986). To this end, the director often recruits detailed descriptions of the object (e.g., "it looks like a triangle shape on top" rather than shorter descriptions like "pyramid"). Yet it is only after the matcher has provided positive evidence of understanding (e.g., "ok I got it") or demonstrated understanding ("oh yeah like a pyramid") that the referential expression has been grounded (Clark & Brennan, 1991). After this grounding process, the pairs over multiple referrals start to devise more economical and efficient references to the objects, either due to natural proclivities of efficient communication and/or due to pressures to complete the game in a fast way. The challenge here is to construct an economical and portable reference that somehow allows the pair to easily identify the referent, overcoming limitations of memory, ambiguity, semantic transparency, and likely other communication-typical bottlenecks. This second challenge is often met by using references that derive from the initially grounded referential expressions (e.g., "pyramid"), thus leading to alignment (Brennan & Clark, 1996; Pickering & Garrod, 2004).

The detailed process and consequences of the pair settling or choosing a certain portable referential expression to be used throughout future interactions are not well-studied and certainly not often studied in a way that is computationally reproducible and automatable. This is an important gap in the literature, as it is precisely this settling or sifting process, where references become entrained through reuse in interaction, that is likely an important aspect of how languages evolve (to this day) and how conventionalization unfolds. Being able to automatically track the re-use of referential expressions in natural conversation promises to provide a novel lens through which to advance our understanding of these core-linguistic dynamics of reuse.

Methodologically, several issues arise for tracking joint reuse. All types of words are reused all the time in conversation (for the same language is spoken), and it is not immediately clear how to decide what shared vs. non-shared references are. This issue needs to be addressed by finding the right level of abstraction and constructing the right selection criteria for the discovery of relevant reused material within 'messy' conversation. There have been several pro-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

posals for automatically tracking different types of linguistic reuse, including syntactic and lexical levels of representation (Dideriksen et al., 2023; Duran et al., 2019; Fernández & Grimm, 2014; Healey et al., 2014; Nenkova et al., 2008). Below, we present what we believe is a computationally effective way to automate the process of discovering *shared constructions* where participants align in form and meaning. With these methodology in hand, we address the following key questions. Broadly, we aim to understand when and how often shared constructions arise and whether such constructions appear to mediate the outcomes of the interaction (i.e., the establishment of naming conventions as measured after social interaction). We conjecture that share constructions provide a useful level of analysis to investigate the dynamics of reference negotiation in interaction, and hypothesize that their properties and patterns of use are directly associated with the level of convergence participants exhibit after having interacted.

To answer these questions, we analyse a corpus of 66 dialogues where participants with director-matcher roles engage in a repeated referential task with novel objects, and also perform an individual naming task for each novel object before and after the interaction. Using this data and our method to automatically detect shared constructions, we conduct three analyses. Analysis 1 examines how often shared constructions arise and the degree of variation within the shared constructions that emerge for a referent. We put our method to the test by devising a strong baseline using pseudo-pair conversations. These pseudo-pairs help us differentiate between shared constructions that might simply arise because the objects invite certain ways of referring to them versus shared constructions that arise because of the pair-unique conversational history that is built over time. Analysis 2 studies the relationship between shared constructions and individual speaker labels for objects before and after the interaction. Finally, Analysis 3 investigates whether the features of shared constructions can explain the speakers' convergence when labelling objects following their interaction.

Together, these analyses shed new light on the dynamics of cross-speaker linguistic convergence, using an automated method suitable for quantifying convergence in large dialogue corpora.

Methods

Dataset

Our study utilises transcribed conversations from 66 pairs of Dutch-speaking participants. Concretely, the data we use comes from two multimodal datasets: 19 dyads are part of the dataset introduced by Rasenberg et al. (2022), and 47 dyads are part of the CABB dataset (Eijk et al., 2022), but participants in both datasets completed the same referential director-matcher task. Overall, the combined corpus includes 27 all-female, 10 all-male, and 29 mixed-gender dyads, with participant ages ranging from 18 to 33 years old (22.6 average). Participants were not acquainted with each other beforehand.

In this dataset, participants alternate between director and matcher roles to jointly identify images of 16 novel 3D objects called "fribbles" (see Figure 1), first proposed by Barry et al. (2014). They do so for a total of six rounds. In each round, the director describes a fribble for the matcher to identify among the 16 candidates. The 16 fribbles are displayed on a screen (one screen per participant, visually inaccessible to the respective other) and are distributed randomly over 16 positions in each round. The performance of the pairs in resolving references is near the ceiling (the mean accuracy is 99.4%). On average, the interactive task takes 25 minutes.

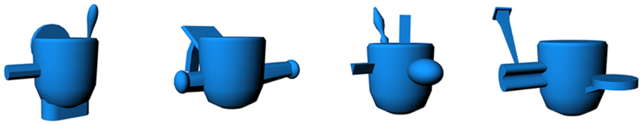

Figure 1: Four example fribbles used as stimuli in the task.

Before and after the collaborative, interactive task, speakers participate in an individual naming task, where they are asked to name each fribble with 1 to 3 words that would allow their partner to identify it amongst the other fribbles. These names are not produced within dialogue interaction and hence reflect the lexical labels that the individual participants associate with the objects before and after the communicative task.

Extracting Shared Lemmatised Constructions

We are interested in capturing the reuse of linguistic behaviour across dialogue partners when they refer to the same object; hence in instances including alignment of both form and meaning (Rasenberg et al., 2020). Regarding form, we focus on cross-speaker matching of lemmatised speech, i.e., speech where content words (nouns, adjectives, verbs, and adverbs) are reduced to their base form. For example, the Dutch word "ball" (singular noun), "ballen" (plural noun) and "balletje" (noun with diminutive suffix) can all be reduced to the lemma "ball" (ball). This captures a level of representation that is more abstract than lexical alignment with exact word matching while still being semantic (in contrast to syntactic alignment). We believe that cross-speaker matches at the level of lemmatised content word constructions that refer to the same object can be good candidates for capturing 'conceptual pacts' (Brennan & Clark, 1996), i.e., dialogue-based agreements about how to refer to an object.

Inspired by the approach of Sinclair and Fernández (2021), who (unlike us) focus on purely form-based lexical and syntactic alignment in the context of language acquisition, we propose a method to automatically identify the alignment of lemmatised constructions with a common referent. We adapt their computational method to extract matching sequences from dialogue transcripts to our data and purposes as follows.

Using the Python library spaCy (Honnibal et al., 2020), we first remove disfluencies, tag each word with its part of speech category, and lemmatise it to its base form. Next, for

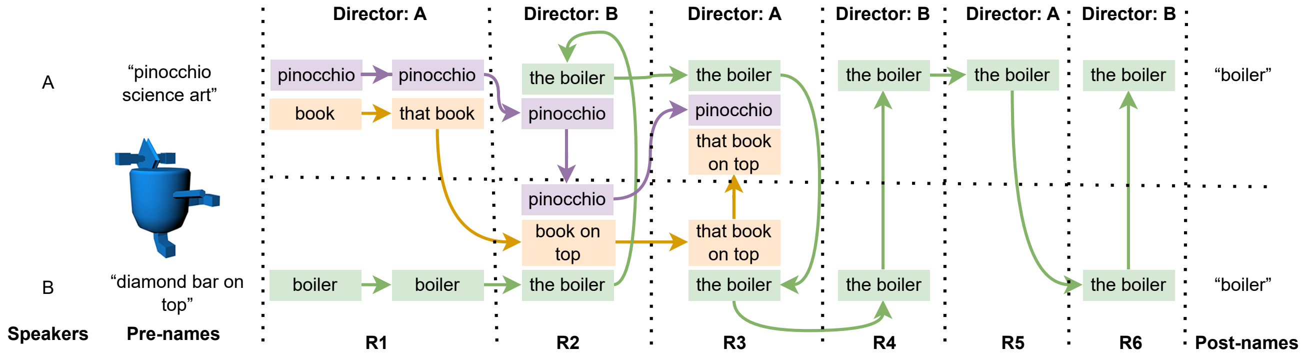

Figure 2: This figure shows the pre- and post-interaction names and the shared construction types for a fribble by a pair. Before the interaction, speakers A and B refer to the fribble as "pinocchio science art" and "diamond bar on top," respectively. The figure shows the shared constructions that emerge, with the arrows indicating the order in which speakers repeat these constructions. For instance, in the first round, speaker A, acting as the director, refers to the fribble as "pinocchio" twice, and speaker B repeats this construction in the second round. This dyad uses three shared construction types for this fribble (indicated by the colours purple, orange, and green). The types "book" and "pinocchio" are dropped after Round 3, while the type "boiler" is used in all rounds (a total of 11 times) as well as in the post-interaction names by both speakers.

each dyad and each fribble, we combine all trials of the dialogue (across rounds) where the participants are trying to identify that fribble. Using the sequential pattern-matching algorithm proposed by Dubuisson Duplessis et al. (2021), we extract all the sequences of lemmas (including single lemmas) that both dialogue participants have used across these sections of the dialogue. Finally, we filter out the sequences consisting exclusively of function words as well as those that are used for multiple fribbles, because they tend to correspond to generic phrases used across the board (e.g., "the main part" or "head").[1].

This procedure results in a set of *shared lemmatised constructions* (or *shared constructions* for short) per dyad and fribble. These constructions can be further grouped into ***shared construction types*** based on their common content lemmas. For instance, in the example shown in Figure 2, we group the shared constructions "dat boek bovenop (that book on top)", "boek bovenop (book on top)", "dat boek (that book)", and "boek (book)" uttered by both participants of a dyad to refer to a given fribble into the shared construction type "book". As a result, our approach detects shared constructions with a common lexical core per referent throughout the entire dialogue, in contrast to other automated methods such as the ALIGN package (Duran et al., 2019), which captures linguistic alignment as turn-by-turn overlap.

Pseudo-Pairs

To establish a baseline for our study, we create a dataset of 66 control dialogues constructed by combining speech by two participants from two different dialogues within the corpus, respecting the same director-matcher roles as in the original dialogues and using speech where the two). combined speakers refer to the same object. However, unlike actual participant pairs, these pseudo-pairs did not directly interact in the referential communication task. Pseudo-pairs allow us to control for "shared" constructions that may arise by chance (e.g., because all participants refer to the same set of objects) rather than as a result of interaction.

Computation of Lexical Cosine Similarity

To measure the degree of form-based similarity between the names that the speakers produce in the individual naming task and between these names and shared lemmatised constructions, we use the same method as Rasenberg et al. (2022)—lexical cosine similarity. To measure the similarity between two expressions, we compute the cosine similarity between the corresponding expression vectors with binary values (1/0), indicating whether a given lemma is present. Effectively, this measure captures lemma overlap while controlling for expression length, resulting in a similarity score ranging from 0 (no similarity) to 1 (full overlap). For example, the cosine similarity between "pinocchio nose above" and "window pinocchio nose" is 0.67. Naming cosine similarity will become relevant in Analyses 2 & 3.

Analysis 1: Presence of Shared Constructions and their Patterns of Use over Dialogue Rounds

We start by quantifying the degree of linguistic alignment and how this evolves over a dialogue. Using the method described above, we automatically extract all the shared constructions and shared construction types for all dyads and all fribbles over all the six rounds of each dialogue. We do this for the 66 dialogues that make up the corpus as well as for the pseudo-pairs. This analysis provides information about the dynamics of shared constructions as well as a confirmation that shared constructions emerge in conversation (rather than simply be-

[1] Our code and data are available on GitHub at https://github.com/EsamGhaleb/SharedLinguisticConstructions

ing invited by the object's characteristics).

We find that 92% of participant pairs (61 out of 66) use at least one shared construction for each fribble. Furthermore, an average of 34% of all utterances per dialogue include shared constructions. Figure 3 shows that this rate increases as the interaction progresses: from 27% in the first round to 37% of utterances containing shared constructions in the final round (Spearman's $\rho = 0.36, p \ll 0.001$). In contrast, only 14% of utterances contain shared constructions in the pseudo-dialogues, with no increase over rounds. This indicates that alignment is largely the result of interaction, rather than solely the consequence of all dyads referring to the same objects.

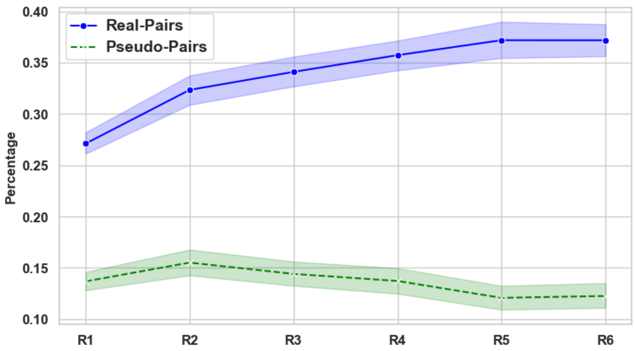

Figure 3: Percentage of utterances containing shared constructions over the rounds of interaction.

Participant pairs tend to use several shared constructions per fribble. We find that, on average, dyads use 4 shared construction types per fribble. Different ways of referring to the same object may arise because participants refer to parts of a fribble—for example, as illustrated in Figure 2, they may refer to the top part of the displayed fribble as "book"—or they may arise because participants entertain different conceptualisations, e.g., using "pinocchio" vs. "boiler" to holistically refer to the object as a whole. Either way, participants tend to converge on a smaller number of shared construction types towards the end of the dialogue (see, e.g., Figure 2, where only "boiler" is used for this fribble in the last three rounds of the task). We find that the average number of shared construction types per fribble in the first round is 4.25, significantly higher than the average of 1.86 present in the last round ($t = 16.45, p \ll 0.001$).

Overall, from this initial analysis we conclude that (1) our method to automatically identify shared constructions captures interaction-driven linguistic alignment, as shown by the contrast in the rate of shared constructions in real-pairs vs. pseudo-pairs; (2) in line with findings in earlier work, linguistic alignment is pervasive in the communicative interactive task, with shared constructions emerging for almost all fribbles and being present in 34% of utterances; (3) participants initially use several shared constructions types per referent and over time some of these constructions tend to be dropped, as common ground is built up.

Analysis 2: Individual Speaker Names versus Interactive Shared Constructions

In our second analysis, we investigate the relation between cross-speaker alignment in the interactive task, as captured by shared constructions, and the names given to the fribbles by each participant in the individual naming task before and after the interaction. This analysis provides key information about what pairs bring to the table before the interaction and whether shared constructions established in interaction are linked to referential labels used after the dialogue.

We start by examining whether the way in which participants individually name the fribbles is altered after the interactive task. For this, we calculate cosine similarity between the pre- and post-interaction names given by a participant for each fribble. If a speaker used the same name before and after the dialogue, we would obtain a cosine similarity of 1. Instead, we find that the average cosine similarity is 0.27 (std = 0.24), indicating that participants tend to name fribbles differently after the communicative task.

To investigate to what extent this difference is mediated by interaction-based linguistic alignment, we compute the cosine similarity of the dyads' shared construction types for a fribble with the pre- and post-interaction names of each participant. On average, $41.3\% \pm 0.09$ of pre-interaction names and $61.5\% \pm 0.10$ of post-interaction names per participant overlap with the shared constructions. As shown in Figure 4, we find that shared constructions are more similar to post- than to pre-interaction names. The plot also shows that the similarity of shared constructions with post-interaction names is higher for shared constructions used in the later rounds of the dialogue (Spearman's $\rho = 0.2, p \ll 0.001$). Besides this recency effect, we also observe a frequency effect: the more frequently a shared construction is used by a participant, the more similar their post-interaction name will be to that construction (Spearman's $\rho = 0.45, p \ll 0.001$).

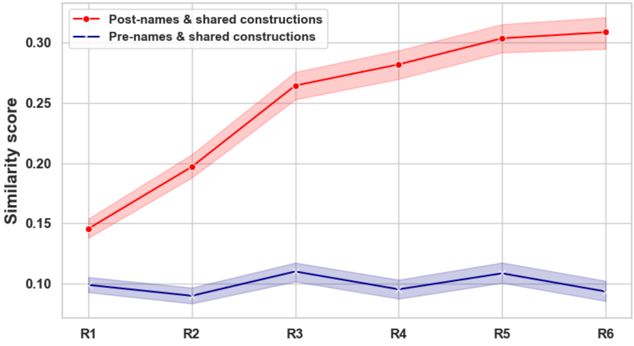

Figure 4: Lexical cosine similarity between a speaker's pre- and post-interaction names and the shared constructions used for a fribble, over the six dialogue rounds.

Taken together, these results suggest that referential labels used individually after the interaction are strongly associated with shared constructions forged in interaction.

Analysis 3: Naming Convergence Across Speakers is Linked to Shared Constructions

Finally, we investigate how similar the names individually given by each speaker to the fribbles before and after the interaction are to each other. Rasenberg et al. (2022) and Eijk et al. (2022) showed that the level of cross-speaker name similarity increases after the communicative task. Here we ask: To what extent is this converging trend related to the patterns of use of shared constructions?

To address this question, we first compute the cosine similarity between the two pre-interaction names (S_{pre}) and between the two post-interaction names (S_{post}) given by the participants and then measure the difference between these two similarities ($S_{post} - S_{pre}$). For actual participant pairs, we find that the mean cosine similarity between the two pre-interaction names is 0.06, while after the interaction the participant names have a mean cosine similarity of 0.43. Thus, there is an average increase of 0.37 cosine similarity. In contrast, for pseudo-pairs of participants, the average cosine similarity between the two post-interaction names is 0.07, and the similarity difference between pre- and post-interaction names is 0 in this case. Hence, there is no increase of similarity in the pseudo-pairs. These results confirm that the converging trend in the post-interaction names already observed by Rasenberg et al. (2022) and Eijk et al. (2022) emanates from the interactive processes taking place during the communicative task, which are absent in the pseudo-pairs.

Recall that our first analysis revealed that speakers may not immediately converge on a unique, simple shared construction type. Instead, they may use several complementary descriptions and entertain different alternative views of a referent, as illustrated by the example in Figure 2. We hypothesise that using many different shared construction types may be indicative of difficulty building common ground and finding a simple way to refer to an object, which could lead to less similar post-interaction names. Indeed, we find a weak negative correlation between the number of shared construction types and the cosine similarity of the post-interaction names (Spearman's $\rho = -0.13, p \ll 0.001$): the more construction types for a fribble, the less similar the post-interaction names tend to be, as shown in Figure 5.

Leaving aside the possible presence of various shared construction types per referent, we analyse two features of the most dominant shared construction type per fribble within a dialogue (e.g., "boiler" in Figure 2): its frequency, i.e., in how many dialogue rounds it is used, and its recency, i.e., how close in terms of rounds the construction is to the end of the dialogue (and hence to the post-interaction naming task). We find that frequency positively correlates with the degree of cosine similarity between the participants' post-interaction names (Spearman's $\rho = 0.28, p \ll 0.001$). That is, the more rounds the dominant shared construction is used in, the more similar the two post-interaction names are to each other. A weaker recency effect is also present: the later in the dialogue the dominant shared construction is used, the higher the degree of post-interaction similarity (Spearman's $\rho = 0.17, p \ll 0.001$).

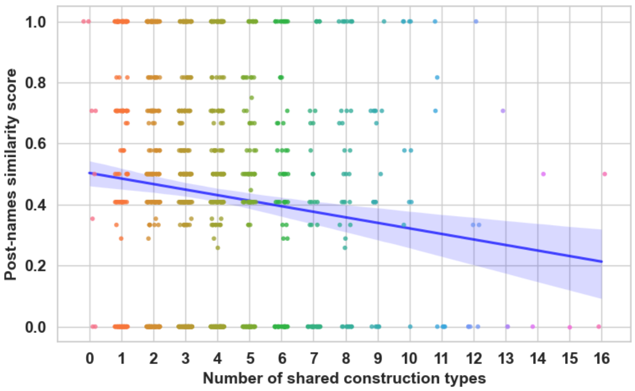

Figure 5: Correlation between the number of shared construction types per object and the cosine similarity between the post-interaction names of the two participants in each dyad.

In summary, this analysis shows that the degree of convergence between speakers post-interaction is mediated by the patterns of reuse captured by shared constructions. More construction types are associated with less convergence, while the frequency and recency of the most prominent type correlate with higher cross-speaker naming similarity.

Discussion

We study the reuse of linguistic behaviours in referential communication through the lens of automatically detected shared lemmatised constructions (or shared constructions for short), further grouped into shared construction types based on their common content lemmas. Our work examines the patterns of shared constructions and studies whether they mediate speakers' convergence in naming conventions for novel referents after social interaction.

Dynamics of Shared Constructions and Cross-Speaker Convergence

First, the study replicates the earlier findings that linguistic alignment is prevalent in referential communication, and we show this is not simply attributable to objects soliciting similar expressions as pseudo-pairs do not linguistically align in this way. The analysis also shows that speakers tend to entertain several shared construction types and eventually converge to fewer ones over the course of the conversation, thus discontinuing the use of aligned-upon content words. This finding goes beyond previous studies which have shown that speakers often simplify referential expressions by dropping hedges (e.g., "like" or "sort of") (Carroll, 1980; Clark & Wilkes-Gibbs, 1986) and closed-class parts of speech (e.g., pronouns, conjunctions and determiners) (Hawkins et al., 2020). Reducing the number of content-word-based shared constructions over the interaction might be linked to establishing a stronger common ground. The third analysis reveals that using a higher number of shared construction types could hinder the establishment of a single, straightforward naming convention. This is evident by the negative relationship between the number of shared construction types and speakers' post-interaction naming similarity. Nevertheless, when considering the most prominent shared construction type, higher usage frequency correlates with higher cross-speaker naming similarity. Thus, the degree of cross-speaker conventionalisation post-interaction is affected by a complex interplay between different properties of shared constructions.

Individual Conventionalisation through Shared Constructions

After the interaction, individual speakers tend to use labels that are strongly related to shared construction types that emerge during the interaction. Our findings indicate that shared construction types become increasingly similar to speakers' post-interaction names in later rounds of the interaction, showing a recency effect. We also observe a strong frequency effect, with individual post-interaction names exhibiting higher similarity with more frequent shared constructions. Despite the observed low similarity between pre-interaction names and shared constructions, pre-interaction names may still be of importance for the conventionalisation process—a preliminary analysis not reported in this paper suggests that pre-interaction names are more similar to shared constructions types than to utterances that do not contain shared constructions. Future work could further investigate how pre-conceptions undergo joint transformation through interaction.

Automatic Detection of Shared Constructions in Referential Communication

We present a novel methodology that automatically captures linguistic alignment, complementing manual approaches used in previous studies such as those by Brennan and Clark (1996) and Rasenberg et al. (2022), which are hard to scale to representative samples. For example, due to the time-consuming nature of manual coding, the study by Rasenberg et al. (2020) limited their focus to the emergence of linguistic alignment in the first round of the dialogue. In contrast, this study offers methods to facilitate the detection of the emergence *and* dynamics of linguistic alignment over time.

Although the methodology presented in this paper provides a valuable tool for studying the reuse of linguistic behaviours in referential communication, it comes with a few limitations. Firstly, it relies on manually transcribed conversations, which may be affected by spelling errors or transcript variations (which are limited in the CABB corpus by using strict transcription protocols and a minimal number of transcribers). Second, one of the caveats in our study is that fribbles comprise a base plus 3 to 6 sub-parts. We observe that speakers commonly start out by describing various sub-parts, but over time they focus on the most salient ones. This might explain why they gradually drop shared construction types and converge to fewer ones as the interaction progresses. This observation also raises a question about whether speakers' shared constructions tend to become more holistic in nature by the end of the interaction (e.g., describing the whole of base plus sub-parts as "boiler"). However, we lack information regarding the specific subpart(s) for which shared constructions are employed. Consequently, we cannot distinguish between cases where participants use a variety of holistic labels or multiple labels for different subparts. These are unanswered questions that can be explored in future studies.

Finally, an interesting future direction could be to explore whether the dynamics of the captured linguistic alignment during the conversation can account for changes in individual conceptualisations of fribbles (e.g., people might come to "see" a fribble as a robot as a result of the conversational history). To this end, behavioural and neural metrics of pre- to post-interaction change in conceptual representations of fribbles made available by Eijk et al. (2022) could be exploited.

Conclusion

This paper investigates how shared linguistic behaviours influence speakers' convergence on labels for novel objects after social interaction. We present an automated method to detect these behaviours, i.e., shared lemmatised constructions, and apply it to a referential communication corpus. The study shows a prevalence of shared constructions in the interactive communicative task. Particularly, speakers use a higher number of shared construction types in initial rounds, which decreases significantly as the interaction progresses. As discussed above, our findings show a negative correlation between the number of shared construction types per referent and the similarity of speakers' post-interaction names. We thus find that more alignment is not always better when it comes to referring to novel objects; aligning on many referential expressions may, in fact, hinder speakers' convergence on an economic naming convention. Instead, we find that using fewer shared construction types (and using those frequently and towards the end of the interaction) is more likely to result in convergence on an object label, as measured post-interaction. The current approach to shared referential construction promises to help accelerate our understanding of how a common ground is carved out of joint action. This common ground building is ultimately related to how meaning is collaboratively negotiated in interaction from the bottom up. In related work (Akamine et al., 2024), we make further progress in this direction by exploiting the present approach to analyse the interplay between lexical and gestural alignment.

Acknowledgments

This research was funded by the Dutch Research Council (NWO) under a Gravitation grant (024.001.006) awarded to the Language in Interaction consortium. Raquel Fernández and Judith Holler are supported by the European Research Council (ERC grant agreement numbers 819455 and 773079, respectively).

References

Akamine, S., Ghaleb, E., Rasenberg, M., Fernández, R., Meyer, A., & Özyürek, A. (2024). Speakers align both their gestures and words not only to establish but also to maintain reference to create shared labels for novel objects in interaction. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*.

Barry, T. J., Griffith, J. W., De Rossi, S., & Hermans, D. (2014). Meet the Fribbles: Novel stimuli for use within behavioural research. *Frontiers in Psychology*, *5*, 103.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition*, *22*(6), 1482.

Carroll, J. M. (1980). Naming and describing in social communication. *Language and Speech*, *23*(4), 309–322.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Clark, H. H., & Brennan, S. E. (1991). *Grounding in communication.* American Psychological Association.

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

Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2023). Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General*, *152*(3), 864.

Dubuisson Duplessis, G., Langlet, C., Clavel, C., & Landragin, F. (2021). Towards alignment strategies in human-agent interactions based on measures of lexical repetitions. *Language Resources and Evaluation*, *55*, 353–388.

Duran, N. D., Paxton, A., & Fusaroli, R. (2019). ALIGN: Analyzing linguistic interactions with generalizable techNiques—A Python library. *Psychological methods*, *24*(4), 419.

Eijk, L., Rasenberg, M., Arnese, F., Blokpoel, M., Dingemanse, M., Doeller, C. F., Ernestus, M., Holler, J., Milivojevic, B., Özyürek, A., Pouw, W., van Rooij, I., Schriefers, H., Toni, I., Trujillo, J., & Bögels, S. (2022). The CABB dataset: A multimodal corpus of communicative interactions for behavioural and neural analyses. *NeuroImage*, *264*, 119734. https://doi.org/10.1016/j.neuroimage.2022.119734

Fay, N., Walker, B., Swoboda, N., & Garrod, S. (2018). How to create shared symbols. *Cognitive Science*, *42*, 241–269.

Fernández, R., & Grimm, R. (2014). Quantifying categorical and conceptual convergence in child-adult dialogue. *Proceedings of the annual meeting of the cognitive science society*, *36*(36).

Hawkins, R. D., Frank, M. C., & Goodman, N. D. (2020). Characterizing the dynamics of learning in repeated reference games. *Cognitive Science*, *44*(6), e12845.

Healey, P. G., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PloS one*, *9*(6), e98598.

Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength natural language processing in Python.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of personality and social psychology*, *4*(3), 343.

Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, *192*, 103964.

Nenkova, A., Gravano, A., & Hirschberg, J. (2008). High frequency word entrainment in spoken dialogue. *Proceedings of ACL-08: HLT, Short Papers*, 169–172.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169–190.

Rasenberg, M., Özyürek, A., Bögels, S., & Dingemanse, M. (2022). The primacy of multimodal alignment in converging on shared symbols for novel referents. *Discourse Processes*, *59*(3), 209–236.

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in multimodal interaction: An integrative framework. *Cognitive Science*, *44*(11), e12911.

Sinclair, A., & Fernández, R. (2021). Construction coordination in first and second language acquisition. *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue*.