COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 47 (2023) e13387

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13387

Language-Specific Constraints on Conversation: Evidence from Danish and Norwegian

Christina Dideriksen,[a] Morten H. Christiansen,[a,b,d] Mark Dingemanse,[c] Malte Højmark-Bertelsen,[f] Christer Johansson,[e] Kristian Tylén,[a,b] Riccardo Fusaroli[a,b,g]

[a]*School of Communication and Culture, Aarhus University*
[b]*The Interacting Minds Center, Aarhus University*
[c]*Centre for Language Studies, Radboud University*
[d]*Department of Psychology, Cornell University*
[e]*Department of Linguistic, Literary and Aesthetic Studies, University of Bergen*
[f]*KMD A/S*
[g]*Linguistic Data Consortium, University of Pennsylvania*

Received 7 February 2023; received in revised form 30 October 2023; accepted 13 November 2023

Abstract

Establishing and maintaining mutual understanding in everyday conversations is crucial. To do so, people employ a variety of conversational devices, such as backchannels, repair, and linguistic entrainment. Here, we explore whether the use of conversational devices might be influenced by cross-linguistic differences in the speakers' native language, comparing two matched languages—Danish and Norwegian—differing primarily in their sound structure, with Danish being more opaque, that is, less acoustically distinguished. Across systematically manipulated conversational contexts, we find that processes supporting mutual understanding in conversations vary with external constraints: across different contexts and, crucially, across languages. In accord with our predictions, linguistic entrainment was overall higher in Danish than in Norwegian, while backchannels and repairs presented a more nuanced pattern. These findings are compatible with the hypothesis that native speakers of Danish may compensate for its opaque sound structure by adopting a top-down strategy of building more conversational redundancy through entrainment, which also might reduce the need for repairs. These results suggest that linguistic differences might be met by systematic changes in language processing and use. This paves the way for further cross-linguistic investigations and critical assessment of the

Correspondence should be sent to Riccardo Fusaroli, Linguistics, Cognitive Science and Semiotics, Aarhus University, Jens Chr Skous vej 2, 8000 Aarhus, Denmark. E-mail: fusaroli@gmail.com

interplay between cultural and linguistic factors on the one hand and conversational dynamics on the other.

Keywords: Conversational dynamics; Common Ground; Interactive Alignment; Backchannels; Conversational repair

1. Introduction

Conversation is the foundation for human social interactions. Through conversations we establish and maintain relationships, share information, and jointly tackle complex interpersonal tasks such as coordinating the pick-up of children from daycare or deciding on a new corporate strategy. This coordination happens through a dynamic flow of words and gestures, building, monitoring, and maintaining mutual understanding and shared knowledge relevant to the situation. And despite its complexity, this collaborative process often appears effortless (Christiansen & Chater, 2022; Fusaroli & Tylén, 2012; Fusaroli, Gangopadhyay, & Tylén, 2014; Garrod & Pickering, 2004).

The dynamics of mutual understanding are an integral part of how we understand human cognition in context (e.g., Dale, Fusaroli, Duran, & Richardson, 2013; Fusaroli, Demuru, & Borghi, 2012; Schilbach et al., 2013), and are becoming crucial components in the development of human–computer interfaces (Loth, Jettka, Giuliani, & De Ruiter, 2015; Sugiyama, Meguro, Yoshikawa, & Yamato, 2018), and clinical assessments of social functioning (Dwyer, David, McCarthy, McKenna, & Peters, 2019; Hopkins, Yuill, & Keller, 2016; Schilbach et al., 2013). However, the extent to which they vary and adapt to different contextual constraints (e.g., a casual chat vs. reaching a joint decision at work vs. helping a child in developing their language) is only now starting to be systematically investigated (Colman & Healey, 2011; Cox et al., 2022a, 2022b; Dideriksen, Christiansen, Tylén, Dingemanse, & Fusaroli, 2023; Healey, Purver, & Howes, 2014). Similarly, while it is acknowledged that cross-linguistic and cross-cultural variations play a crucial role in human behavior and cognition (Christiansen, Contreras Kallens, & Trecca, 2022; Cox et al., 2022a, 2022b; Henrich, Heine, & Norenzayan, 2010; Trecca, Tylén, Højen, & Christiansen, 2021), very little is known about how conversational dynamics systematically vary and adapt to diverse languages and cultures.

The current study addresses exactly these issues: It investigates how conversational dynamics vary across a theoretically interesting linguistic contrast—Danish and Norwegian—and different contextual demands—affiliative (AC) and task-oriented (TOC) conversations. This provides a framework within which to approach "language as a complex adaptive system": how diverse concomitant external constraints shape conversational practices in the short term through changes in task-related constraints and in the long term through differences in the processing affordances of different languages (Beckner et al., 2009).

1.1. Building mutual understanding

The construction of mutual understanding has often been described as a *grounding* process, in which interlocutors explicitly share positive and negative evidence of understanding to build *common ground* (Brennan & Clark, 1996; Clark, 2009; Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). An alternative prominent account—*interactive alignment*—considers the building and maintaining of mutual understanding as a more implicit process where interlocutors prime each other to re-use one another's linguistic structures (Pickering & Garrod, 2004, 2021). No matter the mechanisms involved, mutual understanding is argued to require several components: the entrainment of mental representations, the updating of mutual understanding with new information, and the continuous checks and repairs of the shared understanding. A range of subtle but pervasive conversational devices are argued to facilitate such processes, among which we focus on three specific types of verbal cues: positive evidence of understanding in the form of backchannels (e.g., *uh-huh*), negative evidence in the form of requests for conversational repair (e.g., *what?*), and the repetition of linguistic structures (e.g., the interpersonal entrainment of lexical or syntactic elements).

How and why differences in the use of conversational devices emerge to accommodate understanding across contexts, languages, and cultures is a key outstanding issue. To address this issue, we adopt the wider perspective of language as a complex adaptive system, suggesting that language evolution, change, acquisition, and use are all affected by interdependent linguistic, physical, biological, cognitive, and cultural patterns (Beckner et al., 2009; Christiansen & Chater, 2016; Tylén, Fusaroli, Bundgaard, & Østergaard, 2013). Accordingly, conversations can be conceived as adaptively developing and changing to meet the needs of the activities at hand, as well as cognitive, linguistic, and cultural constraints (Christiansen & Chater, 2022). By conceiving conversations as adaptive in nature, the scientific focus is directed at the different constraints that shape language use at interacting timescales. For instance, contextual demands on the precision of reference can rapidly change as we move from conversation to conversation. And while language-specific and cultural differences (and their affordances for precise information sharing) may be more stable, they might nonetheless also impose equally important constraints. This perspective promises to provide a comprehensive framework for assessing existing findings and lead to new testable hypotheses about what affects mutual understanding in conversation as well as further theory development (as we detail below).

In the following, we synthesize previous findings on the three key conversational devices at stake—*backchannels, conversational repairs, and entrainment*—and their contextual and cross-linguistic variation. This motivates an empirical study that investigates the interplay of task and language-specific contextual demands on conversational devices: in particular, the relation between the local constraints of the activity at hand (task-oriented vs. affiliative conversations) and the global constraints of cross-linguistic differences (Danish vs. Norwegian, two otherwise similar languages that differ primarily in degrees of sound opacity).

1.2. Devices for building mutual understanding in conversation

Research in conversational devices has been scattered across disciplines, with most studies focusing on a single device observed within a specific corpus, and it is consequently difficult to generalize findings across studies and contexts. Following up on recent efforts to build a more comprehensive perspective (Albert & de Ruiter, 2018; Dideriksen et al., 2023; Dingemanse et al., 2015; Fusaroli et al., 2017; Micklos & Woensdregt, 2023), we provide working definitions of the three conversational devices and their functions across contexts.

1.2.1. Backchannels

Backchannels are subtle cues produced by the listener in a conversation (Bangerter & Clark, 2003; Gardner, 2001; Schegloff, 1982; Yngve, 1970). They often overlap with the speaker's utterance and are not considered an interruption or turn but rather a signal of *positive evidence* that the two interlocutors are on the same page, that is, that they have a mutual understanding. While backchannels include non-verbal forms, for instance, head nodding (Hömke, Holler, & Levinson, 2018), we focus here on verbal cues. Specifically, we focus on vocal backchannels, which can be uttered in the form of short words such as *uh-huh* and *yes* (Benus et al., 2007; Truong & Heylen, 2010).

Backchannels are attested across several languages and contexts as a frequent mechanism that comprises between 20% and 50% of utterances in casual conversations (Dideriksen, Fusaroli, Tylén, Dingemanse, & Christiansen, 2019; Dideriksen et al., 2023; Fusaroli et al., 2017). Backchannels have been argued to have a more affiliative function in conversation due to a lower frequency in task-oriented contexts (Dideriksen et al., 2023) but also to vary in their function, for instance, in the degree to which they display recipiency in contrast to constitute fuller responses (Drummond & Hopper, 1993). Backchannels have been argued to vary across languages and cultures, for instance, with Chinese speakers using more frequent backchanneling than Canadian English speakers (Li, Cui, & Wang, 2010). However, a nuanced analysis of the factors that might motivate such differences is still missing.

1.2.2. Repair

While backchannels indicate a continuing mutual understanding between interlocutors, repairs signal the opposite, namely, providing *negative evidence* of a potential disruption of mutual understanding (Schegloff, Jefferson, & Sacks, 1977). By requesting a repair, the listener indicates that mutual understanding is compromised and needs to be re-established. In this way, the listener can communicate that they misheard something or did not quite understand what was previously said. The form and function of repairs have been widely discussed, and corpus analyses show a connection between the two (e.g., Purver, Ginzburg, & Healey, 2003). Three broad ways of initiating repair have been identified on a scale from general to specific (Dingemanse et al., 2015). All examples are Danish and from the corpus analyzed in the current paper, with minimal edits to make them more clear or concise.

Open requests (see Example 1) refer to entire utterances or general statements rather than specific words. Open requests are often short words, such as *huh?* or *what?*

Example 1. Open requests

A	*Men jeg synes kun der var den her jeg kunne vælge* But I thought I could only choose this one
B	*Hvad?* What?
A	*Der var kun den her jeg kunne vælge. Jeg ved ikke om jeg var sent…* I could only choose this one. I'm not sure if I was late…

Restricted requests (see Example 2) refer to a specific part of a previous utterance that needs repair but are still general in the sense that it does not refer to a specific object, location, or person, for instance, *who? where?* or *which?*

Example 2. Restricted requests

A	*Nå, der har jeg en parkeret lastbil* Oh, I have a parked van there
B	Hvor? Where?
A	*Ja helt oppe nordpå* Well, all the way up North

Restricted offers (see Example 3) are specific suggestions for what the listener might have said or intended to say.

Example 3. Restricted offers

A	*Fra fra runestenene altså sådan lidt højere oppe end der hvor der står guldmine?* From from the runic stones, like just a little higher up than where it says goldmine
B	*Okay hvad sagde du det var? En kirkegård?* Okay what did you say it was? A cemetery?
A	*En kirkegård, ja* A cemetery, yes

Repairs have been observed across a wide variety of languages and contexts. Based on a large study of 12 typologically different languages, Dingemanse et al. (2015) found that (1) they all had a system of signaling problems with understanding the content of a previous utterance, (2) they all displayed a preference for more specific types of repair when possible, and (3) people produced one repair per 1.4 min on average. However, this apparent universality of repair does not necessarily mean that no cross-linguistic variation is observed. Indeed, the study, which relied exclusively on naturally occurring affiliative conversations, found a difference in frequency of repair of about a whole minute between the language with the highest repair rate (Argentine Sign Language) and that with the lowest (Dutch). Given the heterogeneity of the corpora analyzed in terms of recording conditions, ongoing activities, settings, and so forth, it is difficult to univocally attribute these differences in frequency to linguistic and typological features of the specific languages. Indeed, three studies have compared affiliative and task-oriented conversations in the lab and in more naturalistic contexts and found that repair happens at much higher frequencies in task-oriented conversations (Colman & Healey, 2011; Dideriksen et al., 2023; Fusaroli et al., 2017). This suggests that more attention should be paid to systematic variations in repair frequencies both within and across languages.

1.2.3. Linguistic entrainment

Entrainment is defined as the repetition of structures, for instance, lexical and syntactic structures between interlocutors in a conversation, but it also includes prosodic, acoustic, and visual aspects such as body posture, facial expression, and accent (Chartrand & Bargh, 1999; Chartrand & Lakin, 2013; Duran, Paxton, & Fusaroli, 2019; Healey et al., 2014; Louwerse, Dale, Bard, & Jeuniaux, 2012; Ostrand & Chodroff, 2021; Pickering & Garrod, 2004; Rasenberg et al., 2020, 2020; Scheflen, 1964; Shockley, Santana, & Fowler, 2003; Wynn & Borrie, 2022). While acknowledging the importance of all these aspects, we focus here on verbal entrainment as a starting point. More specifically, we concentrate on interlocutors' repetitions of each other's lexical, syntactic, and semantic structures during conversation. We illustrate these three types of entrainment with examples below.

Example 4. Lexical entrainment

A	***Jeg har*** *ikke nogen altså jeg har en stor* ***klippe*** *men den ligger* **I** do not **have** any, I mean, I have a big **rock**, but it's
B	***Jeg har*** *to* ***klipper*** **I have** two **rocks**

In Example 4, the lexical forms "I," "have," and "rock" uttered by speaker A are repeated by speaker B constituting a case of lexical entrainment. Note that we consider repetitions at the level of lemmas, that is, "rock" and "rocks" are considered two cases of the same lemma "rock." Also note that some approaches to lexical entrainment focus on repetitions of open class words (e.g., nouns, adjectives, and verbs), while others focus on closed class words (e.g., pronouns, see Niederhoffer & Pennebaker, 2002, for a discussion). We evaluate all words in our approach and control for baseline entrainment—that is, "coincidental" entrainment due to high-frequency words—using surrogate pairs (see Methods section and Supporting Information).

Example 5. Syntactic entrainment

A	*Det er faktisk slet ikke godt* **PRON V** ADV **ADV ADV ADJ** That's actually not very good
B	*Det er overhovedet ikke sundt* **PRON V ADV ADV ADJ** That's not healthy at all

Syntactic entrainment is illustrated in Example 5. Here, two sequences of pronouns (PRON), verbs (V), adverbs (ADV), and adjectives (ADJ) are repeated in the following sentence ("PRON V" and "ADV ADV ADJ"), while the specific words change.

Example 6. Semantic entrainment

A	*Den dér er anderledes end det første* ***rumvæsen*** That one is different from the first **alien**
B	*Det ville sikkert være nemmere hvis vi havde et* ***rumskib*** It would probably be easier if we had a **spaceship**

In Example 6, the two interlocutors, while not entraining lexically, use words that are semantically related and therefore entrain on their topic, namely, "space."

Verbal entrainment has been attested across several languages (Bertrand & Espesser, 2017; Dideriksen et al., 2023; Reitter & Moore, 2014), and it is generally considered widespread in conversations, both in its *rate* of occurrence (how many utterances contain any verbal entrainment) and in its *level* (how much of a given utterance consists of linguistic forms repeated from the interlocutor's previous turn). However, given the heterogeneity of methods employed to measure entrainment, it is hard to compare frequencies across previous studies (Dideriksen et al., 2023; Duran et al., 2019; Rasenberg et al., 2020). Here, we look at entrainment in adjacent turns (imposing a temporal and sequential limit on what counts as entrainment), and we consider lexical, syntactic, and semantic entrainment separately.

While no explicit cross-linguistic comparison of verbal entrainment has been conducted—although see Levitan, Beňuš, Gravano, and Hirschberg (2015) for acoustic-prosodic entrainment—it has been shown to vary across contexts. For instance, differences in the use of conversational devices between affiliative conversations and task-oriented conversations have been reported (Colman & Healey, 2011; Fusaroli et al., 2017; Healey et al., 2014), with findings showing that while instances of entrainment were more frequent in affiliative conversations, more linguistic items and structures were repeated in each instance of entrainment in task-oriented conversations (Dideriksen et al., 2023).

1.3. Conversational devices and varying contextual demands

While there have been persistent efforts to describe the occurrence of individual conversational devices, the research has been disparate thus far, with separate accounts for each mechanism. Moreover, each study typically has different objectives, making it difficult to compare the occurrence and function of the individual devices, as well as their potential interdependencies across conversational contexts. However, there are reasons to think that the devices are interdependent. Schegloff (1982) suggested that because backchannels provide positive evidence of mutual understanding, they also signal indirectly that no repair is needed. Further work has provided conceptual elaboration and empirical evidence for that suggestion (Healey, Mills, Eshghi, & Howes, 2018; Howes & Eshghi, 2021). Even more crucially, the devices can overlap: directly repeating one's interlocutor's words might signal that mutual understanding is in place; but specific repairs tend also to involve the re-use of words to request their clarification (for an exploratory analysis of this, see Fusaroli et al., 2017). Such overlaps in device use are of considerable interest, but only exploratory analyses are performed in this study and should be further targeted in future work.

In particular, one previous study integrated backchannels, repair, and entrainment within a coherent framework (Dideriksen et al., 2023). The study investigated contextual differences in the use of the three conversational devices and their relation to performance in joint tasks. It found consistent patterns across contexts, suggesting that the use of conversational devices in affiliative conversations was of the more generic kinds (e.g., backchannels and open requests), providing less specific support to the speaker or requesting more general clarifications. In contrast, task-oriented conversations featured more specific conversational devices such as restricted offers along with a less frequent but more specific use of entrainment (level of entrainment) possibly due to the enhanced demand for precision in these types of contexts. Thus, the use of conversational devices appeared to be dynamically adjusted to the activity at hand contingent on its varying requirements of precise information sharing and building of rapport (Fusaroli, Rączaszek-Leonardi, & Tylén, 2014).

Evidence for contextual variations in the use of conversational devices in conversation supports arguments for the adaptive nature of these variations. Backchannels and repairs have been conceptually linked to the ability to navigate and jointly solve complex conversational tasks (Bangerter & Clark, 2003; Clark & Brennan, 1991; Mills, Groningen, & Redeker, 2017). Further, linguistic entrainment has been more directly associated with performance, both conceptually and experimentally, albeit the results are not fully consistent across studies, perhaps due to methodological differences (Dale et al., 2013; Dideriksen et al., 2023; Duran et al., 2019; Fusaroli et al., 2012; Fusaroli, Bjørndahl, Roepstorff, & Tylén, 2016; Healey et al., 2014; Pickering & Garrod, 2004; Reitter & Moore, 2014; Tylén, Fusaroli, Østergaard, Smith, & Arnoldi, 2023). In other words, when there is a need to communicate more precisely (task-oriented conversations), interlocutors might, for instance, repeat longer stretches of each other's utterances (lexical entrainment level). Since interlocutors who do so more perform better, this variation in lexical entrainment level could be considered adaptive—albeit only provisionally so, given the observational nature of these studies.

Whether cross-linguistic variation can be productively conceived of as contextual variation systematically affecting the use of conversational devices so that they accommodate differences in linguistic structures and/or cultural practices is still an open question. There are accounts of cross-linguistic variation in the use of conversational devices; however, the findings are mixed, most often with a descriptive focus, and difficult to compare with one another. Indeed, it is unclear how much of the variation between, for instance, Japanese and U.S. English should be ascribed to language-related differences, to differences in cultural practices, or to some combination of the two. That is, are the differences in conversational devices in the United States and Japan due to differences in properties of the two languages (in word order, use of honorifics, etc.) and/or due to historical and cultural differences (e.g., individual vs. collective focus)? Comparisons of existing data cannot answer this question because they tend to derive from separate studies that either involve typologically different languages and/or speakers from different cultures.

In the current study, we take a first step toward assessing the impact of native language on the use of conversational devices by leveraging a natural experiment involving Danish and Norwegian—two typologically and genetically similar languages that, except for substantial phonetic differences, are deeply interconnected culturally, historically, and with regard to their linguistic features. By investigating how conversational dynamics vary between these two languages, we can start developing more precise theories of how phonetic differences might affect conversational language use—which will need to be further replicated and assessed in a broader range of languages and cultures. Thus, we can begin to develop a more comprehensive framework for how people build understanding in conversations and the conversational devices that subserve this purpose. Our framework does not only incorporate local, short-term constraints (such as the degree of precision required by the current conversational context) but also more global, long-term constraints (such as how the nature of a given language might shape information sharing due to, for instance, sound opacity or specific grammatical features). Inferring differences in conversational dynamics to be due to different linguistic affordances is of course an abductive process, meaning that alternative explanations are possible (and indeed will be discussed in later sections). However, by motivating the hypotheses in the literature and using the findings to assess and further elaborate the theory, we provide an explicit target for replications, generalization, development, and, crucially, challenges thus facilitating further theory construction.

1.3.1. A natural experiment: The Danish versus Norwegian contrast

Danish and Norwegian (Bokmål) are two very closely related Northern European languages, spoken in countries with a common cultural background (e.g., see Gooskens, 2006). In particular, the history of Denmark and Norway is closely intertwined, and the two countries are historically, socioeconomically, and culturally very similar. Norwegian and Danish are also typologically very similar and share many of the same lexical and morphosyntactic features. However, Danish has developed a peculiarly opaque sound structure, which is not present to the same extent in Norwegian (see Fig. 1 for an example of how this affects speech signals in Danish, compared to Norwegian).

Specifically, Danish has a skewed ratio of vowel-like sounds (vocoids) to consonant-like ones (contoids), causing the undifferentiated acoustic signal displayed in Fig. 1 (panel a), which might make syllables and words harder to demarcate. Both languages exhibit a large number of full vowels (Norwegian has 18, while Danish has 21; Basbøll, 2005; Kristoffersen, 2000), but Danish has been argued to have a large degree of variation in the use of vowels with at least twice as many phonetically distinct vowel qualities (Basbøll, 2005). Moreover, Danish additionally contains no less than 49 diphthongs, compared to six in Norwegian (Grønnum, 1998, 2005; Kristoffersen, 2000). There is also a tendency in spoken Danish to turn some contoids into vocoids, for instance, /b/ is often realized as [u̯] (as in *løbe* [ˈløːu̯ə], English to run) and /g/ as [ɪ̯] (as in *kage* [ˈkʰæːɪ̯ə], compared to [ˈkʰaːˈkə] in Norwegian, English cake). Last, reduction phenomena such as omission and assimilation of schwa (e.g., *gade* [ˈg̊æːð̞ə] → [ˈg̊æːð̞ð̩], English street) is a very common feature, causing words to become truncated, and thereby blurring the already opaque word boundaries. Note that each of these processes can be found individually in other languages (for instance, consonant lenition is found in

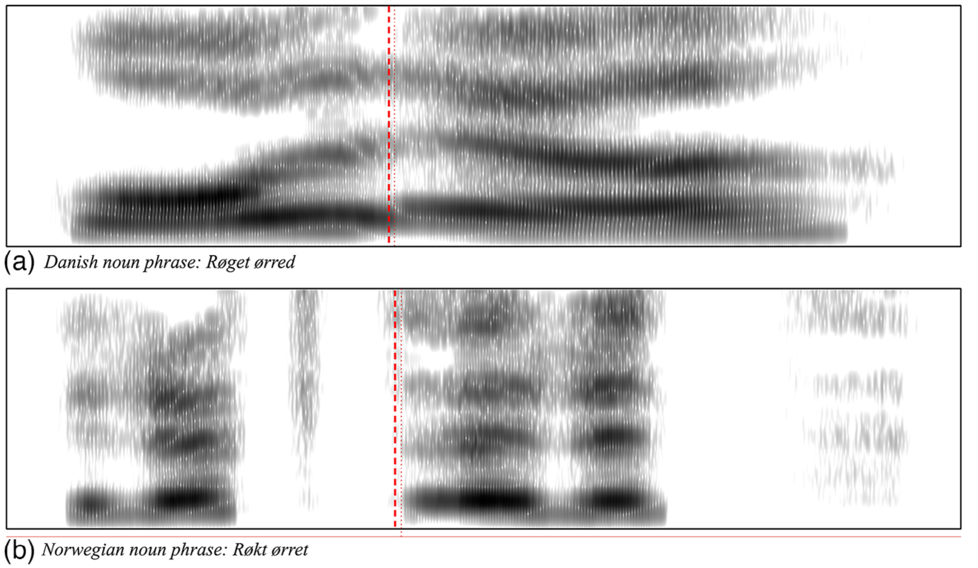

Fig. 1. A spectrogram of "smoked trout" in Danish and Norwegian. The red line represents the approximate word boundary, which is clearly segmented in the Norwegian example but barely detectable in Danish.

Castilian Spanish and schwa assimilation in German; Trecca et al., 2021). However, it is the combination of a large inventory of vowel-like sounds, the tendency to turn some consonants into semi-vowels, and extensive reduction that makes the Danish speech signal rather opaque (Bleses & Basbøll, 2004). By contrast—a crucial for the current study—this phenomenon is not prevalent in closely related Norwegian, where consonants typically are more clearly pronounced and the vocalic inventory smaller.

As displayed in Fig. 1, the relatively opaque sound structure of Danish is likely to affect the segmentation of word boundaries in the speech signal, and thus make it harder to identify the linguistic components of an utterance from acoustic cues alone. Given how dependent early language acquisition is on segmentation (e.g., Mattys & Jusczyk, 2001; Nazzi, Dilley, Jusczyk, Shattuck-Hufnagel, & Jusczyk, 2005), it is not surprising that a series of studies has found acquisition delays in vocabulary and morphology in Danish children, compared to children acquiring other Scandinavian languages (Bleses et al., 2008). Crucially, two experimental studies, directly manipulating the amount of reduction present in the linguistic stimuli, showed that Danish toddlers have a harder time processing vocoid-rich, consonant-reduced utterances, compared to unreduced, contoid-rich ones (Trecca, Bleses, Madsen, & Christiansen, 2018; Trecca et al., 2021). Danish children eventually catch up with vocabulary and morphology acquisition (Bleses et al., 2008), but it appears that speakers require more exposure to and engagement with the language to learn how to process it proficiently (Trecca et al., 2021).

In adulthood, Danes do not have any apparent communicative and linguistic problems. Nonetheless, it seems that diverse strategies are at play; Danish speech might still present a potentially challenging sound opacity, but speakers learn to rely on diverse processing and production strategies to overcome this challenge. Indeed, two recent studies have suggested that native adult speakers of Danish seem to rely more on the context and less on the actual speech signal during speech processing, compared to Norwegian native speakers (Ishkhanyan et al., 2020; Trecca et al., 2021). This indicates that the sound structure of Danish has implications beyond language acquisition and well into adult language use (see Trecca et al., 2021, for an extensive review of Danish language acquisition and processing). A cross-linguistic comparative study of turn-taking in polar question contexts indicates that Danish responses were somewhat slower than most other languages, with an average turn-to-turn response time of 469 ms, compared to, for instance, 7 ms in Japanese (the average across all languages was 208 ms; Stivers et al., 2009). Note, however, that the medians were closer suggesting that more than a generalized slowness of Danish conversations, we might observe a higher frequency of substantial pauses than in other languages, suggestively due to the relative sound opacity of Danish (Trecca et al., 2021).

Existing studies of Danish have mostly focused on compensatory strategies when processing the signal at the level of individual listeners (Ishkhanyan et al., 2020; Trecca et al., 2021). Considering the opaque sound structure of Danish and the recent findings that native speakers of Danish are more reliant on contextual cues, the question arises as to whether this might affect how Danish speakers establish and maintain mutual understanding in conversational interaction. Do native speakers of Danish orchestrate mechanisms of mutual understanding differently, perhaps by generating a more informative context to facilitate processing, compared to speakers of Norwegian? For instance, if Danish is harder to understand, at least in some contexts, higher entrainment might enable interlocutors to effortlessly build mutual understanding by creating a stronger more redundant common ground (both in terms of shared representations but also to the extent that the speaker's repetition of the interlocutor's words makes them easier for the interlocutor to parse) and by enabling a quicker identification of mishearing and misinterpretation (if the repeated linguistic forms are not those uttered by the first interlocutor). This would resonate with the idea that several mechanisms for processing and interactions are possible but that the specific features of a particular language might afford a more productive use of some mechanisms, compared to others. Different languages have different features, which pose slightly different challenges and affordances to the cognitive system and to conversational dynamics. By focusing on sound opacity in Danish and the challenges it seems to pose, we provide an example of this rich interplay between language, cognition, and conversational dynamics, which should be further and more systematically studied.

1.4. Summary and hypotheses

The differences in sound structure between the two otherwise very similar languages—Danish and Norwegian—provide a natural quasi-experiment to explore how the interplay of local and global contextual factors affect the use of conversational devices. To the extent that conversational devices are crucial for building understanding and affiliation in the context of conversational interaction, we expect that Danish-speaking interlocutors might compensate for the more opaque linguistic signal by employing conversational devices more frequently than Norwegian-speaking interlocutors. Specifically, we hypothesize (H1) that speakers of Danish, a language that is arguably harder to process than Norwegian (Trecca et al., 2021), will present an overall higher frequency of conversational devices (across the three devices described in Section 1.2), to compensate for the opaque sound structure. Note that since all devices have been associated with the establishment and maintenance of mutual understanding, all could be involved. Therefore, one contribution of the current study is to provide empirical data for more precise theoretical claims about how conversational dynamics might be related to sound opacity. Further, we expect (H2) to see an interplay between conversational demands, where differences in conversational devices between task-oriented conversations and affiliative conversations will be moderated by the increased effort by Danish speakers to monitor, build, and repair mutual understanding in conversation. If, for instance, we observe the lexical entrainment rate to be lower in task-oriented conversations than in affiliative conversations, we would still expect such difference to be relatively smaller in Danish than Norwegian since lexical entrainment rate might also be required to ensure an easier processing of the Danish speech signal, and therefore cannot be excessively reduced in task-oriented conversations. Note that the direction of the interaction is not part of the hypothesis but is only presented as an example.

2. Methods

2.1. Corpora and participants

We elicited conversations from 160 participants divided into 80 dyads (40 Danish and 40 Norwegian). Alignment, repairs, and backchannels in the Danish corpus have been previously analyzed in Dideriksen et al. (2019, 2023), and our focus here is their contrast with the Norwegian data. Three dyads were excluded due to technical issues, leaving a total dataset of 77 dyads. Each dyad participated in four sessions: two affiliative conversations and two task-oriented conversations. Participants were asked to complete a questionnaire prior to the conversations, stating age (Danish corpus: mean 23.2, $SD = 3.5$; Norwegian corpus: mean = 23, $SD = 3.4$), gender (Danish corpus: 58% female; Norwegian corpus 51% female), education (Danish corpus: five participants had finished a high school degree, and 75 participants were university students; Norwegian corpus: one participant had finished a high school degree, and the remaining 79 either had a degree from a university or were university students), and dialectal affiliation (Danish corpus: Aarhusiansk: 38% Københavnsk: 6%, Vestjysk: 6%, other: 50%; Norwegian corpus: Østlandsk 35%, Bergensk 20%, Vestlands 13.75%, other 31.25%). The Danish participants were recruited through the SONA recruitment system hosted by the Cognition and Behavior lab at Aarhus University and consisted mainly of university students. The Norwegian participants were recruited through student groups and by handing out flyers on the campus of the University of Bergen. Conversations took place within the respective university campuses, in quiet rooms. Both Danish and Norwegian participants went through the same experimental procedure.

For the first affiliative conversation, we provided the participants with a sheet of open-ended conversation starters (e.g., “Discuss and agree on two superpowers that you would like to have,” see Supporting Information for the full list of conversation starters) and asked them to get acquainted for a while.

Afterward, they were asked to complete the two collaborative tasks. First, they played the Alien Game (Tylén et al., 2023), a joint decision task, where two participants have to jointly make decisions about how to categorize items (in this case, a series of aliens) presented on a shared screen. Through repeated trials, an alien would appear on the screen for 3 s after which it would disappear. This was done to make sure that the participants would have to rely on their memory of the stimulus during the verbal interaction. If they categorized the alien correctly, they would receive 100 points. If they failed, they would receive a penalty of −100 points. The game terminated after 10 min, yielding a variable number of trials depending on the pace and collaborative style of the participants.

Participants then had to solve the Map Task (Anderson et al., 1991; Fay et al., 2018). The task is an asymmetric game, where participants take turns giving directions to one another as to how to draw a path on a map, with the matcher being free to interact, ask for explanations, and so forth. The participants were assigned their own monitor and were separated by a screen that blocked the view to the partner’s monitor but still enabled interlocutors to see each other’s facial expressions. The participants switched roles after each map so that each participant in turn acted as director and matcher. Again, the game terminated after 10 min, yielding a variable number of maps solved by each pair.

For the second affiliative conversation, if no conversation arose spontaneously, the participants were instructed to continue discussing the conversation starters provided for the first affiliative conversation. Usually, however, the participants naturally continued talking about the games without any need for the experimenter to prompt them. Each conversation lasted for about 10 min.

The Norwegian corpus consists of 38 dyads, 150 conversations, and a total of 37,569 utterances (interpausal units surrounded by silence longer than 1 s or interrupted by the other interlocutor) organized into 17,890 conversational turns (sequences of speech by one interlocutor without interruption from the other), consisting of 181,982 transcribed words. The first affiliative conversation had an average of 156 turns per conversation (12.6 words per turn) and the second of 146 (12.9 words per turn). The first task-oriented conversation (the Alien Game) had an average of 229 turns per conversation (6.68 words per turn) and the second (the Map Task) of 165 turns per conversation (9.85 words per turn).

The Danish corpus consists of 39 dyads, 153 conversations, and a total of 38,259 utterances organized in 27,510 conversational turns, consisting of 318,999 transcribed words. The first affiliative conversation had an average of 157 turns per conversation (14.5 words per turn) and the second of 143 (15.1 words per turn). The first task-oriented conversation (the Alien Game) had an average of 233 utterances per conversation (8.21 words per turn) and the second (the Map Task) of 184 (10.8 words per turn).

Danish interlocutors were not reliably better than Norwegian interlocutors in the Alien Game (difference in points: 224, 95% Cis −664, 1058; BF = 2), and only slightly better in the Map Task (difference in standardized deviation from the original path −0.15, 95% Cis 0.03, 11; BF = 10.7).

All conversations were transcribed orthographically by 39 native speakers of the respective language, using parallel channels by interlocutor, and time-coded. As some of the analyses required a more linear presentation of the transcription to identify successive turns by different interlocutors, conversational turns were identified as sequences of speech by one interlocutor without interruption from the other, thus generating an ABABAB structure. Fully overlapping speech between interlocutors occurred occasionally (e.g., during some backchannels) and was interleaved at the first pause longer than 1 s in the speaker holding the floor. The impact of different definitions of conversational turns on, for example, measures of alignment has been assessed in (Fusaroli, Weed, Rocca, Fein, & Naigles, 2023a, 2023b) and showed the robustness of the findings to changes in definition. Relying on videos and transcripts, backchannels and the three kinds of repairs were manually coded. Coding schemes can be retrieved here: https://osf.io/x3s6w/?view_only=05b55fde57a1432f976d328fac6eec98. The first author trained and supervised each coder, presenting examples (see coding scheme), manually checking with them the first 20 min of coded materials, and resolving doubts and difficult cases as the coding proceeded. Additionally, four conversation sets were coded by two independent raters to assess the reliability of the coding schemes. Intercoder reliability was moderate to substantial (Backchannels: kappa = 0.62; Repairs: kappa = 0.65). All coders and transcribers were blind to the hypotheses. Linguistic entrainment was automatically measured (see Statistical Modeling section).

2.2. *Equipment*

Conversations were recorded with a GoPro Hero 5 camera, and sound was recorded with an omni-directional microphone connected to the camera. The stimulus presentation and response collections were set up in Psychopy3 (Peirce & MacAskill, 2018).

2.3. *Statistical modeling*

To test whether there were cross-linguistic differences in the use of backchannels, we built binomial regression models with a Bernoulli likelihood and a logit link. The context (affiliative vs. task-oriented) of the conversation and the language spoken (Danish vs. Norwegian) predicted the presence or absence of backchannel for any given turn. Individual and pair-level propensities to use backchannel were accounted for using varying intercepts (also called random intercepts) for the rate parameter. A Bayesian multilevel modeling approach was chosen, as it allowed for a much more flexible framework to implement and assess complex likelihood functions better reflecting the generating processes behind the data. Further, a Bayesian approach enabled us to comparatively assess the impact of using diverse priors (e.g., based on previous literature or conservatively skeptical of any difference due to language and context). More details on implementation, such as priors, prior impact assessment, and other quality checks are reported in the Supporting Information (e.g., see Figs. S1–S11).

Estimates from the model are reported as mean and 95% compatibility intervals (CI) of the posterior estimates. We test our hypotheses on the full models, that is, whether the frequency of backchannels differs between languages (e.g., more backchannels in Danish than Norwegian) and whether language affects the effects of conversational contexts (e.g., bigger increase in backchannels for affiliative conversations vs. task-oriented conversations in Danish, compared to Norwegian). We calculate evidence ratios (ERs) in the form of the posterior probability of the directed hypothesis against the posterior probability of all the alternatives; that is, if we expected increased backchannels in Danish, we count the posterior samples compatible with this hypothesis and divide them by the number of posterior samples compatible with a null or negative effect. We also report the credibility of the estimated parameter distribution; that is, the probability that the true parameter value is above 0 if the mean estimate is positive, or below 0 if it is negative. When our hypotheses are not adequately supported by the data (ER below 3) or when we hypothesized no difference, we also test for evidence in favor of the null.

To determine whether repair is modulated by language and conversational demands, we used the same procedure as for backchannels but with the absence or presence of repair as the predicted outcome. To test the more specific hypothesis that repair type would be affected by language and conversational context, we include only utterances that were coded as repair and repeat the same procedure as for backchannels with the specific type of repair as the predicted outcome.

To test whether linguistic entrainment is modulated by contextual demands, we first quantify the degree of turn-to-turn lexical, syntactic, and semantic entrainment between speakers. Lexical entrainment is based on lemmatized words; that is, we reduce all inflected forms (e.g., "dogs" or "are") to their dictionary form or lemma (e.g., "dog" or "be"), relying on the Danish and Norwegian Bokmål language models on UDPipe (Straka, Hajic, & Straková, 2016; Strømberg-Derczynski, 2018).

Syntactic entrainment relies on bigram Parts-Of-Speech (POS) tagging, which has been shown to be a good proxy for the entrainment of syntactic structure (Reitter & Moore, 2006). Parts-of-speech tagging relies on the default POS taggers for Danish and Norwegian Bokmål in UDPipe, which have a reported accuracy above 95%, albeit not necessarily on conversational data (Straka et al., 2016; Strømberg-Derczynski, 2018). For each utterance, we produce a list of *n*-grams (2-grams, 3-grams, and 4-grams); that is, for contiguous sequences of parts of speech of length *n* within the same utterance, by inferring parts-of-speech tags (adverb, noun, etc.) from the original (non-lemmatized) transcripts. If an utterance contains fewer parts of speech than necessary to create the *n*-gram, syntactic alignment to that sentence is considered impossible and therefore excluded from the analysis. Since the results are analogous using different *n*-grams, we only report those related to 2-grams, as 3- and 4-grams analyses exclude more utterances due to insufficient sentence length. See Table 1 for an example. Note that this approach deviates from some previous approaches (e.g., Hopkins et al., 2016; Reitter & Moore, 2014) that removed repeated words before calculating syntactic entrainment.

Semantic entrainment is based on FastText Word2Vec representations based on the Danish and Norwegian versions of Wikipedia (Bojanowski, Grave, Joulin, & Mikolov, 2017). Word embeddings encode the meaning of a word as a vector of values. In this way, words that appear in similar contexts (i.e., co-occur with similar words) are closer in the vector space (they have similar values) and thus are assumed to have similar meanings. Word embeddings are widely and effectively employed for text processing purposes such as automated

Table 1
Example of transcript pre-processing

Original	Lemmatized	Parts of Speech	2-Grams
A: She wants to play	[“she”, “want”, “to”, “play”]	[(“she”, “PRP”), (“want”, “VB”), (“to”, “IN”), (“play”, “VB”)]	[(“PRP”, “VB”), (“VB”, “IN”), (“IN”, “VB”)]
B: She gets to play with the bath toys	[“she”, “get”, “to”, “play”, “with”, “the”, “bath”, “toy”]	[(“she”, “PRP”), (“gets”, “VB”), (“to”, “IN”), (“play”, “VB”), (“with”, “PP”), (“the”, “DT”), (“bath”, “NN”), (“toys”, “NN”)]	[(“PRP”, “VB”), (“VB”, “IN”), (“IN”, “VB”), (“VB”, “PP”), (“PP”, “DT”), (“DT”, “NN”), (“NN”, “NN”)]

Note. Parts of speech reported here are abbreviated: Personal Pronouns (PRP), verb (VB), infinitive marker (IN), determiner (DT), noun (NN), preposition (PP).

translation. While the discursive genre of conversations arguably is not akin to that of online encyclopedias, there are no sufficiently large-scale corpora of conversational Danish and Norwegian available to create reliable word embeddings representations from scratch (Kirkedal, Plank, Derczynski, & Schluter, 2019; Strømberg-Derczynski et al., 2020). Each word was associated with the 300 values identifying its position in the 300-dimensional vector space of the word, and we average word embeddings within each utterance. Note that to better account for potentially atypical contextual use of words in our conversational corpora and assess the robustness of our methods, we also generate contextualized word embeddings based on a multilingual transformer model including Danish and Norwegian (see Section S1.3 in the Supporting Information) and assess whether the results change. The results display the same qualitative patterns (see Table S1 in the Supporting Information).

To calculate entrainment, we first identify all pairs of successive utterances produced by the two interlocutors and transform them into numerical vectors for lexical, syntactic, and semantic forms. Thus, the lexical vector includes all unique lemmas present in at least one of the utterances in the pair. Each lemma constitutes a “dimension” of the vector, and the number of occurrences of that lemma in a given utterance is the value under that dimension. The syntactic vector includes all unique *n*-grams of parts of speech present in at least one of the utterances in the pair. Each unique *n*-gram constitutes a “dimension” of the vector, and the number of occurrences of that *n*-gram in each utterance is the value under that dimension. The semantic vector is the utterance-level word embeddings representation described above. See Table 2 for an example. If a conversational turn could not be processed for a given measure (e.g., a one-word utterance could not be converted to 2-grams of parts of speech), alignment to and from that turn was considered to be “NA” and therefore excluded from further analysis.

Linguistic entrainment is calculated as cosine similarity (i.e., the cosine of the angle between two vectors) between successive conversational turns according to the following

Table 2
Examples of lexical, syntactic, and semantic vectors

	Lexical Vector	Syntactic Vector	Semantic Vector
	She, want, to, play, get, with, the, bath, toy	(“PRP”, “VB”), (“VB”, “IN”), (“IN”, “VB”), (“VB”, “PP”), (“PP”, “DT”), (“DT”, “NN”), (“NN”, “NN”)	300 dimensions
She wants to play	1, 1, 1, 1, 0, 0, 0, 0, 0	1, 1, 1, 0, 0, 0, 0	0.0232, 0.02515, 0.027185, …
She gets to play with the bath toys	1, 0, 1, 1, 1, 1, 1, 1, 1	1, 1, 1, 1, 1, 1, 1	0.027013, 0.0117625, −0.01897 …

formula:

$$similarity\,(A, B) = \frac{A \cdot B}{||A|| \times ||B||} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}, \tag{1}$$

where A and B represent the vectors, respectively, for the first and the second interlocutor’s utterance, and *i* indicates the *i*th dimension in the vector. Note that because semantic embeddings involve continuous values, they tend to be more similar to each other than is the case for lexical and syntactic vectors, which contain discrete values (i.e., the quantity of units present, with a strong prevalence of 0s and 1s).

There are alternative ways to calculate entrainment in the literature (see Duran et al., 2019, for a review). Some focus only on content words or on selected word classes. Others ignore turn-by-turn dynamics to focus on the overall similarity over longer stretches of conversation. These different choices are likely to lead to different results. Here, we follow the standardized procedure described and motivated in Duran et al. (2019): Turn-by-turn dynamics of entrainment is a better operationalization of the theoretical construct of linguistic entrainment and more akin to the way we analyze backchannels and repairs. Moreover, including all words reflects the notion that function words play important roles in linguistic style and matching thereof (Ireland et al., 2011; Ireland & Henderson, 2014; Tausczik & Pennebaker, 2010).

Entrainment scores are clearly multimodally distributed for lexical and syntactic entrainment with a peak at 0 (no entrainment) and a second peak within the 0 and 1 boundaries Thus, a high number of turns do not align with the previous turn (see Fig. 2). However, when there is any entrainment, the distribution is centered above zero and decreases steadily at both sides (see Fig. 2). Semantic entrainment, on the contrary, presents no zero inflation: No pair of turns is infinitely distant in the continuous semantic space of words or sentence embeddings.

We therefore chose to use zero-inflated beta regressions to model lexical and syntactic entrainment, and a beta regression to model semantic entrainment, all within a Bayesian multilevel framework. By separating the occurrence of no entrainment (zero inflation) from the actual level of entrainment in those utterances where there is at least some entrainment, we

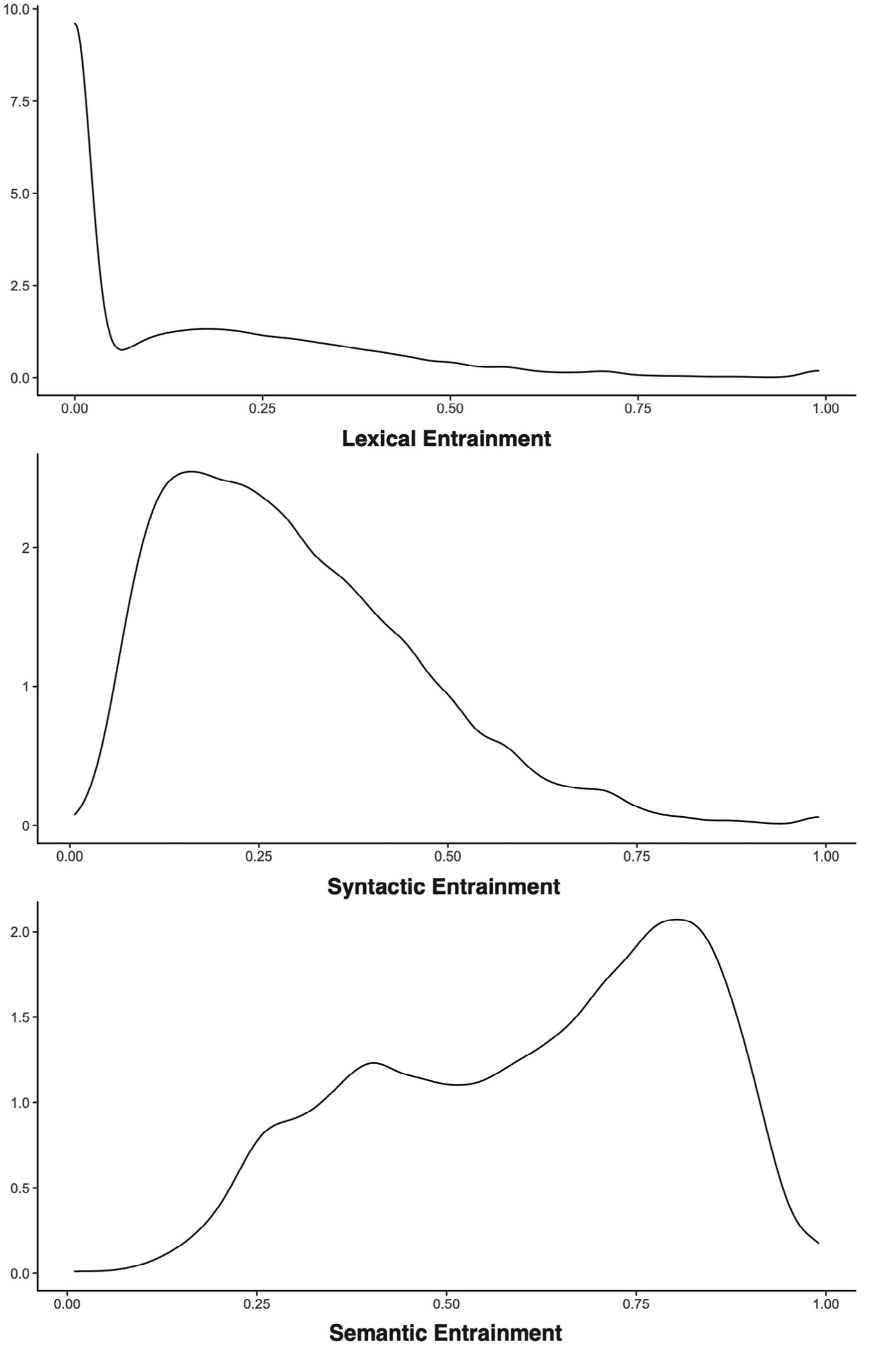

Fig. 2. Density plots of linguistic entrainment. X-axes indicate cosine similarity on a 0–1 scale (with 0 being *no entrainment*, and 1 being *perfect repetition*). Y-axes indicate the probability density estimates for the different values of cosine similarity.

respect the distributional nature of the data, avoiding model misspecifications (e.g., assuming that a mean entrainment would be an adequate description of the bimodal distribution). This distinction also makes sense from a conceptual perspective: We want to know both how relevant it is to entrain at all in a given context (entrainment rate or the inverse of the zero inflation), and how much entrainment there is within an utterance when it is entrained (entrainment level). In other words, we might have situations where we often "touch base" by entraining one or two words (high entrainment rate, low entrainment level) and others where entrainment occurs less often, but once it happens, it involves longer sequences of entrainment (low entrainment rate, high entrainment level). Note that we report entrainment rates as the negative of the inflation term (1 minus the zero-inflation rate), thus indicating the log odds rate of any entrainment, instead of the rate of no entrainment. We report entrainment level as the log odds of the cosine similarity for the utterances containing entrainment. We condition zero inflation, and average level of entrainment on the conversation context, language, and individual and pair-level variability.

Note that concerns have been raised about whether measures of syntactic and semantic entrainment provide any non-overlapping information, compared to lexical entrainment (e.g., Hopkins et al., 2016). Some studies have argued that lexically entrained items should be removed before calculating syntactic and semantic entrainment (Reitter & Moore, 2014). However, that would break the sequence of, for instance, parts of speech, creating non-realistic utterances (see Example 1S in the Supporting Information). Therefore, we provide control analyses in the Supporting Information (see Sections S2.3.2 and S2.3.3), where we stratify the analyses of syntactic and semantic entrainment by lexical entrainment (adding lexical entrainment as a predictor to the model). In other words, we assess syntactic and semantic entrainment beyond lexical entrainment (as if lexical entrainment was set to zero). The control analyses display the same qualitative patterns as the main analyses and corroborate the robustness of the findings (see Tables S2–S5 in the Supporting Information).

There has been some debate on whether this method of measuring entrainment considers the chance level of entrainment adequately. The baseline frequency of different words and parts of speech as well as the constraints of a conversational context might affect the baseline frequencies (Healey et al., 2014). While contrasting between conditions or languages would seem to be immune to these concerns, nevertheless, several confounds could arise. For instance, if one condition is associated with shorter utterances, entrainment measurements might be more extreme (one-word utterances can only have 0 or 1 entrainment), and this could impact the statistical estimates of the difference between conditions. Therefore, in the Supporting Information, we report a control analysis involving surrogate pairs (see Fig. S10, Table S7). We build surrogate pairs by artificially interleaving the utterances of speakers from two different pairs within the same condition. Note that this procedure overcomes the potential bias due to different turn lengths in the two languages, as they are baselined by surrogate pairs within the same language. We, then compare entrainment in real and surrogate pairs via plots and statistical models including an interaction between task and type of pairs. We find that entrainment is credibly higher in real pairs, compared to surrogate pairs for all measures of entrainment. Detailed results are reported in the Supporting Information (see Fig. S10).

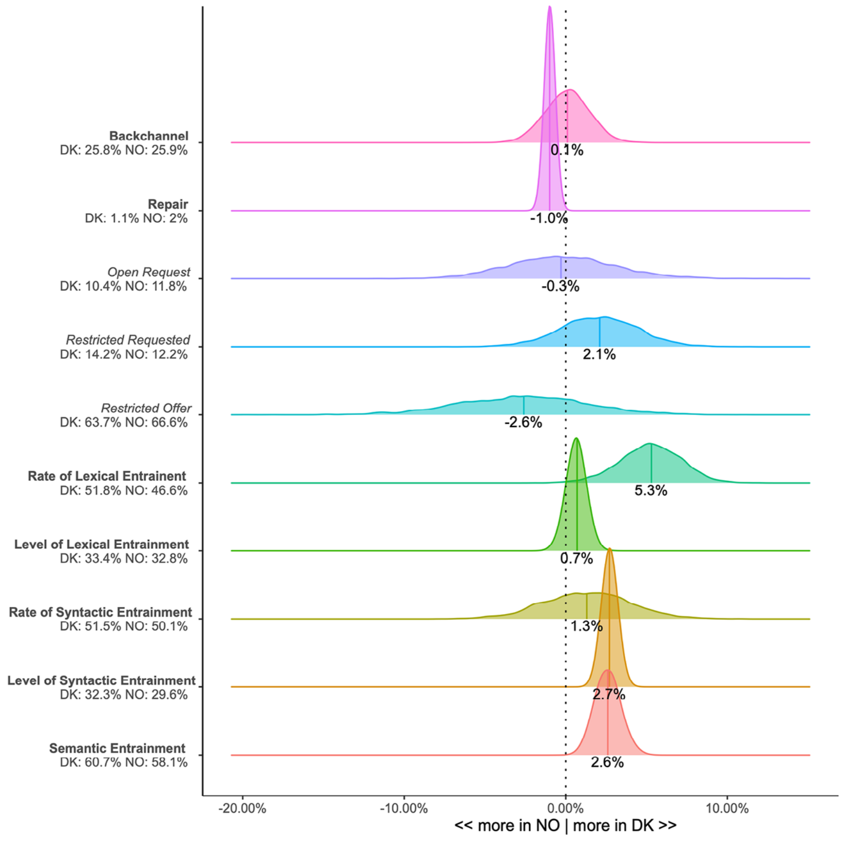

Fig. 3. Effects of language (NO = Norwegian, DK = Danish) on conversational devices across conversational contexts (i.e., averaging across affiliative and task-oriented conversations). Posterior estimates (based on 4000 posterior samples, see Supporting Information for details) for each mechanism on a percentage scale are depicted in the left column, while the plot illustrates the posterior estimates of differences between the two conditions on a percentage scale. Note that backchannel and repair are reported on the total number of utterances, while repair types are reported based on the total number of repair utterances.

2.4. Computational implementation

The implementation relied on R 4.1.1, Rstudio 1.4, tidyverse, brms, cmdstanr, and Stan (Bürkner, 2017; Carpenter et al., 2017; Csardi & Nepusz, 2006).

3. Results

For the full details on the results see Table 3 and Figs. 3 and 4. Of theoretical interest, we found partial support of H1; that is, that Danish involves a higher frequency of conversational devices. Entrainment presents the clearest picture, being generally higher in Danish than Norwegian (except for syntactic rate). Lexical entrainment occurs in 5.3% more of the utterances in Danish (ER = 265.67), and when it occurs, 0.6% more of the lexical choices

Table 3
The frequency of conversational devices between contexts (Affiliative, AC vs. Task Oriented, TOC) and language (Danish, DK vs. Norwegian, NO)

Parameter	AC DK	TOC DK	AC NO	TOC NO	Difference	Interaction
Backchannel	32.95%, $\beta = -0.71$ (-0.84, -0.59)	19.77%, $\beta = -1.4$ (−1.51, −1.29)	31.33%, $\beta = -0.78$ (−0.92, −0.66)	21.18%, $\beta = -1.31$ (−1.43, −1.19)	0.1%, $\beta = -0.01$ (−0.12, 0.11), ER = 1.14	3.1%, $\beta = 0.16$ (−0.01, 0.33), ER = 16.32
Repair	0.68%, $\beta = -4.99$ (−5.24, −4.75)	1.73%, $\beta = -4.04$ (−4.23, −3.86)	1.36%, $\beta = -4.29$ (−4.49, −4.08)	3.03%, $\beta = -3.47$ (−3.65, −3.29)	**1%, $\beta = -0.64$** **(−0.81, −0.46), ER > 1000**	1%, $\beta = -0.13$ (−0.45, 0.19), ER = 3
Open requests	23.4%, $\beta = -1.19$ (−1.78, −0.62)	4.22%, $\beta = -3.12$ (−3.68, −2.62)	22.25%, $\beta = -1.25$ (−1.76, −0.81)	5.87%, $\beta = -2.77$ (−3.22, −2.38)	0.3%, $\beta = -0.14$ (−0.56, 0.27), ER = 2.56	**2.8%, $\beta = 0.41$** **(−0.37, 1.18), ER = 4.14**
Restricted requests	16.67%, $\beta = -1.61$ (−2.01, −1.21)	12.04%, $\beta = -1.99$ (−2.37, −1.6)	14.32%, $\beta = -1.79$ (−2.15, −1.43)	10.31%, $\beta = -2.16$ (−2.55, −1.75)	**2%, $\beta = 0.18$** **(−0.16, 0.51), ER = 4.24**	0.6%, $\beta = 0.01$ (−0.63, 0.63), ER = 1.05
Restricted offers	46.79%, $\beta = -0.13$ (−0.52, 0.26)	77.73%, $\beta = 1.25$ (0.86, 1.65)	49.78%, $\beta = -0.01$ (−0.36, 0.33)	80.06%, $\beta = 1.39$ (1.03, 1.76)	**2.7%, $\beta = -0.13$** **(−0.45, 0.18), ER = 3.02**	0.7%, $\beta = 0.02$ (−0.58, 0.63), ER = 1.07
Lexical entrainment rate	53.78%, $\beta = 0.15$ (0.01, 0.2)	49.9%, $\beta = 0$ (−0.12, 0.1)	51.37%, $\beta = 0.05$ (−0.08, 0.2)	41.82%, $\beta = -0.33$ (−0.44, −0.22)	**5.3%, $\beta = -0.21$** **(−0.33, -0.08), ER = 265.67**	**5.7%, $\beta = 0.23$** **(0.08, 0.38), ER = 113.29**
Lexical entrainment level	32.27%, $\beta = -0.74$ (−0.78, −0.7)	34.59%, $\beta = -0.64$ (−0.68, −0.59)	31.46%, $\beta = -0.78$ (−0.82, −0.74)	34.12%, $\beta = -0.66$ (−0.71, −0.61)	**0.6%, $\beta = 0.03$** **(−0.01, 0.07), ER = 6.65**	2.3%, $\beta = 0.02$ (−0.04, 0.07), ER = 2.1
Syntactic entrainment rate	60.8%, $\beta = 0.44$ (0.13, 0.69)	42.09%, $\beta = -0.32$ (−0.51, −0.12)	60.99%, $\beta = 0.45$ (0.19, 0.69)	39.29%, $\beta = -0.44$ (−0.62, −0.25)	1.3%, $\beta = -0.05$ (−0.24, 0.13), ER = 2.15	**3%, $\beta = 0.12$** **(−0.1, 0.34), ER = 4.77**
Syntactic entrainment level	31.62%, $\beta = -0.77$ (−0.82, −0.72)	33.07%, $\beta = -0.71$ (−0.76, −0.64)	27.9%, $\beta = -0.95$ (−1, −0.9)	31.36%, $\beta = -0.78$ (−0.83, −0.73)	**2.7%, $\beta = 0.13$** **(0.1, 0.16), ER > 1000**	**1.4%, $\beta = 0.1$** **(0.04, 0.16), ER = 265.67**
Semantic entrainment	61.99%, $\beta = 0.49$ (0.43, 0.55)	59.31%, $\beta = 0.38$ (0.33, 0.42)	60.83%, $\beta = 0.44$ (0.38, 0.5)	55.29%, $\beta = 0.21$ (0.16, 0.26)	**2.6%, $\beta = 0.11$** **(0.05, 0.16), ER > 1000**	**2.6%, $\beta = 0.12$** **(0.04, 0.19), ER = 399**

Note. Percentages indicate the proportion of utterances identified as either backchannel, repair, or entrainment (for entrainment rate). Repair types (open request, restricted request, restricted offers) are percentages within the subset of repair utterances only. For entrainment level, we report the percentages of linguistic structures that are repeated from the previous utterance. Beta estimates and 95% compatibility intervals (in parentheses) are reported in the original scale of the model (log odds). “Difference” indicates the average difference between languages across conversational contexts (i.e., averaging across affiliative and task-oriented conversations. “Interaction” indicates the effect of language on the difference between conversational contexts. Credible differences and interactions—for which the evidence ratio is above 3—are in bold. AC, affiliative conversation; TOC, task-oriented conversation; DK, Danish; ER, evidence ratio; NO, Norwegian.

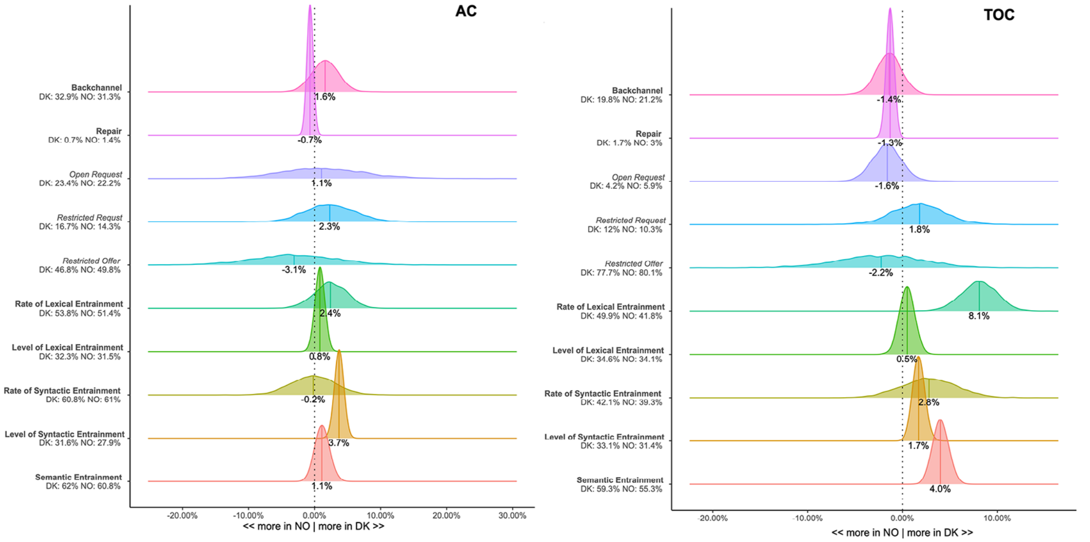

Fig. 4. Effects of language on conversational devices within each of the two contexts. (a) (left) Posterior values of the difference between the two languages in affiliative conversations (AC), while (b) (right) are posterior values of the difference between the two languages in task-oriented conversations (TOC). Posterior values for each mechanism are depicted in the left column, while the plot illustrates the posterior estimates of differences between the two conditions. Note that backchannels and repair are reported relative to the total number of utterances, while repair types are reported relative to repair utterances.

are entrained than in Norwegian (ER = 6.65). Syntactic entrainment shows similar rates between languages (difference 1.3%, ER = 2.15), but more content is entrained in Danish than in Norwegian (2.7%, ER > 1000). Semantic entrainment is also 2.6% higher in Danish (ER > 1000). However, the other devices show a less straightforward picture. There are no reliable differences in the use of backchannels between languages (average difference: 0.1%, ER = 1.14), and repairs are in general more frequent in Norwegian than Danish (average difference of 1% of the total amount of utterances, ER > 1000), with restricted offers being relatively more frequent in Norwegian (2.7% of all repair turns, ER = 3.02) and restricted requests less frequent (2% of all repair turns, ER = 4.24).

We found that H2 (i.e., context-related differences are moderated by the increased effort needed to maintain mutual understanding in Danish due to its opaque sound structure) is compatible with the patterns observed for entrainment but not for other conversational devices. Task-oriented conversations in Danish show a much higher entrainment (except for lexical entrainment level) than in Norwegian. In other words, Norwegian shows a bigger effect of context, with task-oriented conversations displaying a much lower entrainment than affiliative conversations; while in Danish, this effect is smaller: Since all conversations are generally highly entrained, task-oriented conversations are only slightly less entrained than affiliative ones.

Backchannels and repairs do not conform to our hypothesis. Backchannels show a more pronounced effect of conversational context in Danish, with the frequency being slightly higher in affiliative conversations and slightly lower in task-oriented conversations, compared to Norwegian (the effect of conversational context being 3.1% higher in Danish, ER = 16.32). Repairs show the opposite pattern, where Norwegian shows a higher frequency in task-oriented conversations than Danish leading to a slightly higher effect of conversational context in Norwegian (1% higher, ER = 3). Of all the specific types of repair, only open requests show an interaction similar to that observed for backchannels: that is, the frequency is slightly higher in affiliative conversations and slightly lower in task-oriented conversations in Danish than it is in Norwegian, leading to a stronger effect of conversational contexts in Danish.

The cross-linguistic differences are also observed within conversational contexts (see Fig. 4 and Table 3), where entrainment overall is more frequent in Danish speakers, while overall repairs, in particular restricted offers, were more frequent for Norwegian speakers.

To summarize, we find a more complex picture than predicted, where some conversational devices are more prevalent in Danish (such as entrainment), while others are more prevalent in Norwegian (such as repairs). These results will be further explored and interpreted in detail in the discussion.

4. Discussion

4.1. *Overview of the results*

In the current study, we developed a framework in which conversations are construed as adaptive systems, shaped by contextual and linguistic constraints. We relied on a natural quasi-experiment contrasting Danish and Norwegian, two languages with a largely shared cultural and historical background but with a salient difference in sound opacity, to provide a concrete operationalization of the framework and add further details to the theory. In particular, we tested two hypotheses. First, that cross-linguistic differences would affect the use of conversational devices, in particular that the more opaque sound structure of Danish would lead to a higher frequency in the use of conversational devices than in Norwegian to accommodate a more challenging input processing (H1). Second, we predicted that this language effect would interact with the more established findings that conversational contexts affect the use of conversational devices (with an overall higher baseline employment of conversational devices in Danish—due to opaque sound structures—leading to lower differences between task-oriented and affiliative conversations than in Norwegian; H2).

We found that entrainment patterns supported both H1 and H2: Danish native speakers do indeed entrain more on their lexical choices, syntax, and semantic content than Norwegians do—a robust finding across all our control analyses (see Section 2.3 in the Supporting Information). Furthermore, the reduction in entrainment in task-oriented conversations, compared to affiliative conversations, is less marked in Danish, compared to Norwegian, arguably because entrainment is helpful in compensating for the opaque sound structure and cannot therefore be decreased too much.

However, backchannels and repairs showed a more complex picture. Contrary to H1, we found no cross-linguistic differences in the use of backchannels, and overall repair was lower in Danish. Against H2, but in accordance with our main effects of language, we found that the effects of conversational contexts for backchannels and repairs were not smaller in Danish. Indeed, backchannels (which had no overall difference between languages) were actually lower in task-oriented conversations for Danish than for Norwegian. Repairs, which were generally higher in Norwegian, were even higher in the task-oriented conversational context in Norwegian than in Danish (see Example 8).

Example 8. An example of how Danish and Norwegian speakers deal with the complexity of information sharing differently. Lexical entrainment is shown in corresponding pink and green colors. Repair is highlighted in bold. Here, Danish speakers use entrainment to avoid complications, while the Norwegian speakers use repair.

	Danish		Norwegian
A	Det var den der, ikke?	A	Eller også farge, det sk-går jo også an, og hvis den er blå og
B	Ja mm	B	**Kombinasjonen med de?**
A	**Den har ikke pletter**	A	mm
B	**den har ikke pletter det** er næsten det samme som, ja		
A	Ja **det** var **det det** var blå okay måske kan man **ikke** få lov så		
	English translation		**English translation**
A	It's that one, right?	A	Or color also, that- will also do, and if it is blue also
B	Yes uh-huh	B	**The combination of the two?'**
A	**it does not have spots**	A	uh-huh
A	**It does not have spots**, it's almost the same as, yes		
A	Yes **it** was **it it** was blue, okay, maybe you're **not** allowed then		

The results suggest that the strategies involved in anticipating and dealing with the more opaque sound structure of Danish are more nuanced than expected. The initial anticipation of an overall higher frequency of conversational devices in Danish were met for verbal entrainment, where the opaque sound structure seems to motivate a more redundant (entrained) conversation. However, backchannels and repairs were not co-opted in the same way—if anything, the high degree of entrainment might reduce the need for repair. This suggests a cross-linguistic difference in how information sharing in conversations is supported (see Example 1). In other words, if sound opacity is met by increased verbal entrainment in Danish (compared to Norwegian), repairs might be less necessary, and this will result also in a lower increase in repairs when the context requires more precise information sharing. We will now discuss the details of this interpretation for each

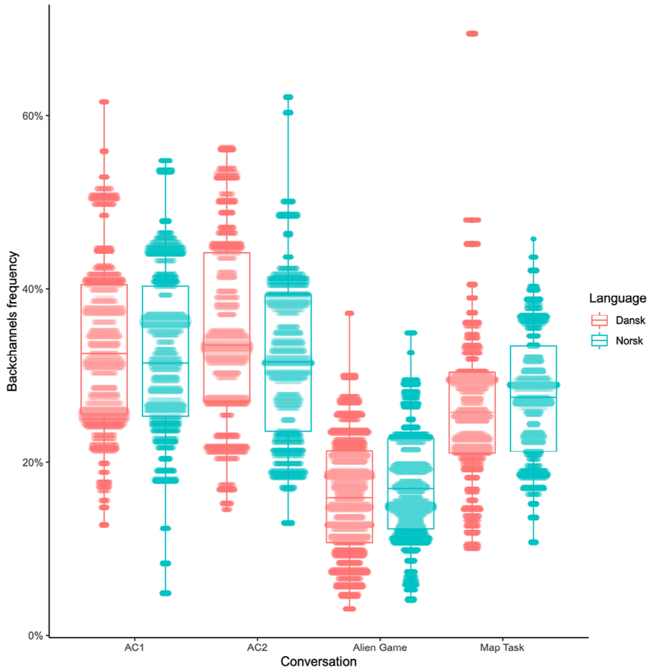

Fig. 5. Frequency of backchannels for each language across conversational contexts. Each datapoint is a conversation within one pair, boxplots overlaid are centered on the median.

conversational device, contextualize it in the perspective of "language as a complex adaptive system," and discuss implications and limitations of the findings.

4.1.1. Backchannels

While there are many descriptions of backchannels and their function, less attention has been dedicated to their frequency of occurrence, making it difficult to compare the findings of this study to the previous literature. We found very similar instances of backchannels in the two corpora, and overall, they were very frequent with one backchannel every 8.5 s (28 words) in affiliative conversations and one every 10.79 s (36 words) in task-oriented conversations in the Danish corpus. Likewise, Norwegian displayed one backchannel every 8.9 s (17 words) in affiliative conversations and one every 10.6 s (21 words) in task-oriented conversations (see Fig. 5). This corresponds to every third turn in affiliative conversations and every fifth turn in task-oriented conversations.

The findings for backchannels thus go contrary to H1. The similarity in occurrence across the two languages strongly suggests that the use of backchannels is not affected by sound structure opacity. Likewise, the results did not show any interaction with conversational demands, indicating that the occurrence of backchannels in conversations with an increased need for precise information sharing were not moderated by linguistic differences.

Nevertheless, we did find the same strong effect of context in both languages, not only between affiliative conversations and task-oriented conversations but also between the two different tasks: affiliative conversations present the highest frequency of backchannels, followed by the Map Task and then by the Alien Game (see Fig. 5). This might be a consequence of the amount of shared information accessible to the participants. In the Alien Game (the condition with the lowest frequency of backchannels), participants share the key information (the alien to categorize) on their screen. This might act as a partially externalized representation of mutual understanding, making continuous confirmation of understanding via backchannels less critical. By contrast, in the Map Task, the two participants have differential access to information, with only the director knowing the path to take. This creates a stronger demand for the listener to backchannel in order to display an understanding of the instructions from the director, thereby confirming mutual understanding (Clark & Brennan, 1991; Clark & Schaefer, 1989; Clark & Wilkes-Gibbs, 1986). More speculatively, the even higher rate of backchannels in affiliative conversations could be due not only to face managing strategies (as suggested in previous literature; Cutrone, 2011) but also because of the high flow of information: Participants tend to shift topic more often than when constrained by a task and their opinions can be more varied. Summing up, the relation of backchannels to affiliative conversations is supported, but we also suggest they are connected to asymmetries in access to information in task-oriented contexts.

4.1.2. Repair

Repairs were less frequent than backchannels in the corpora. Nonetheless, the results showed a higher cross-linguistic difference for repairs, compared to backchannels (see Fig. 6). In the Danish corpus, we found an average occurrence of repair once every 5 min and 4 s (0.2 repairs per minute, one repair every 1013 words) in affiliative conversations, and once every 2 min and 5 s (0.48 repairs per minute, one repair every 417 words) in task-oriented conversations. The Norwegian corpus displayed a higher occurrence of repairs: once per 3 min and 19 s in affiliative conversations (0.3 repairs per minute, one repair every 398 words) and once every 1 min and 15 s in task-oriented conversations (0.8 repairs per minute, one repair every 150 words).

These findings display a clear difference in the use of repairs in Danish and Norwegian but in the opposite direction to what we expected. The overall higher frequency of repairs in Norwegian is consistent across contexts, and apart from restricted requests, we also find higher frequencies across all types of repair, thus disconfirming H1 for repairs (see Fig. 7). However, this linguistic difference is modulated by the use of repair across conversational contexts: As the need for precise information sharing increases (in task-oriented conversations), the number of repairs increase but more so for Norwegian than for Danish. If we infer from the previous results that the increased redundancy in Danish conversations due to linguistic entrainment makes repairs less necessary, then an increased need for precise information sharing due to context might also be largely met through the use of more entrainment and therefore repairs would increase less in Danish than they do in Norwegian.

Across languages, we find that our results replicate previous findings of repair types being modulated by contextual constraints (Dingemanse et al., 2015; Gries, 2005; Reitter & Moore,

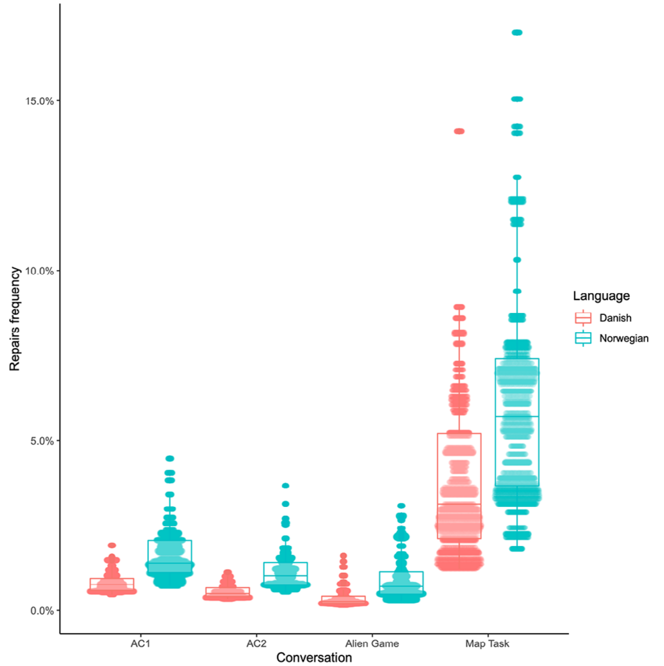

Fig. 6. Frequency of repairs for each language across conversational contexts.

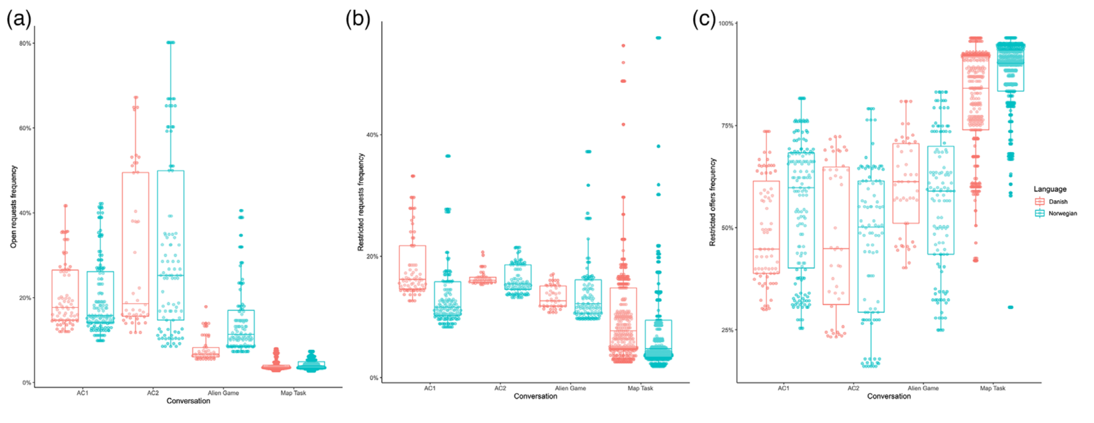

Fig. 7. Frequency of repair types for each language across conversational contexts. (a) Displays open requests, (b) restricted requests, and finally (c) restricted offers.

2006, 2014; Slocombe et al., 2013). Both Danish and Norwegian speakers have a higher frequency of general repairs, such as open requests and restricted requests in affiliative conversations, while task-oriented conversations present a stronger need for the more referentially specific repair type: restricted offer. Furthermore, restricted offers are more prevalent in situations of asymmetric access to information–- the Map Task–- similar to what we observed

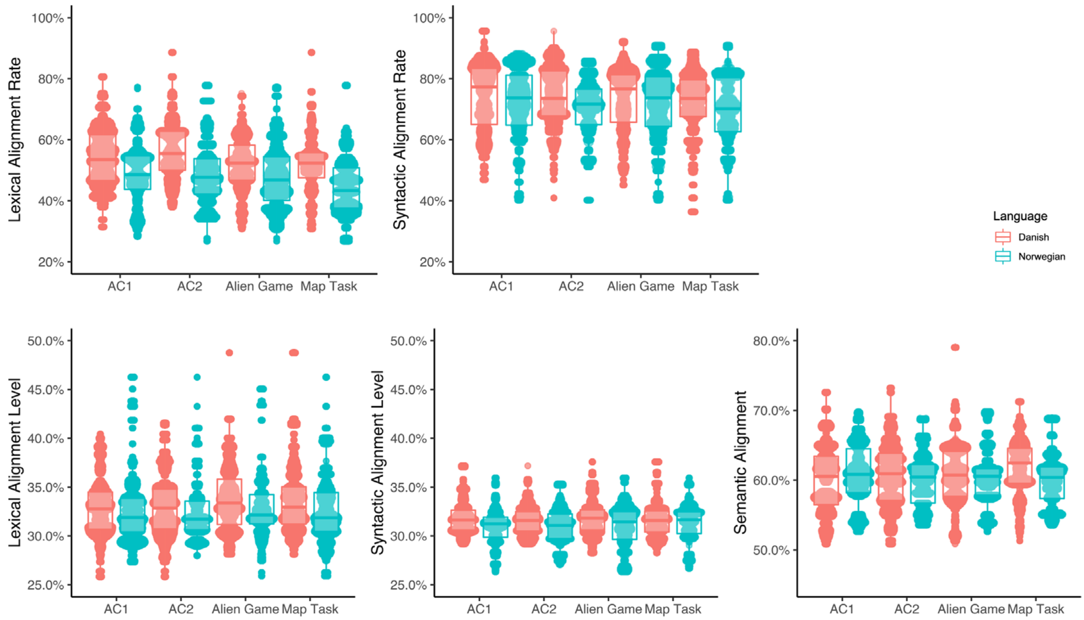

Fig. 8. Rate and level of lexical and syntactic entrainment for each language across conversational context (AC1 = first affiliative conversation, AC2 = second affiliative conversation).

for backchannels. Indeed, they are so prevalent in the Map Task that they drive most of the contextual demand effects of repair use. The frequency of all other forms of repair in task-oriented conversations appears lower than or equal to that in affiliative conversations. These findings further underscore the need for a nuanced articulation of repair types and their function (Dingemanse et al., 2015) and emphasize the importance of considering the impact of context on the use of conversational devices. Moreover, it replicates the finding that specific types of repair—while costly to produce—are preferred in conversation. This is in line with the principle of least collaborative effort, whereby by producing the most specific repair, the listener minimizes the joint effort of the interlocutors (Clark & Wilkes-Gibbs, 1986).

Summing up, the findings are compatible with linguistic constraints affecting the way people deal with problems and breaches in mutual understanding. However, they do so in an unexpected way, with native speakers of Danish relying less on repairs than Norwegian speakers.

4.1.3. Linguistic entrainment

In line with previous studies, we found entrainment to be widespread with more than 45% of utterances including some entrainment (i.e., entrainment rate), across contexts and languages. Likewise, the level of entrainment was also high for both languages, as entrained utterances contained roughly over 30% of repeated forms (see Fig. 8).

The findings—with a single exception—are compatible with both H1 and H2, with native Danish speakers entraining more often—and across more of their words, syntax, and semantics—to their interlocutors than Norwegian speakers, and this consistently interacts with conversational context (e.g., if the context affects the relative entrainment—with lower levels in affiliative conversations—this difference will be larger for Norwegian than Danish). The one exception—syntactic entrainment rate—showed no main effect of language, but it did show an interaction of language and context, with Danish speakers consistently displaying higher rates than Norwegians in task-oriented conversations. These findings are robust even when stratifying syntactic and semantic entrainment by lexical entrainment and using multilingual contextual word embeddings (see Section S2.3.2 and S2.3.3 in the Supporting Information).

As with repair and backchannels, we observe context-driven differences between rate and level, replicating for Norwegian previous findings for Danish (Dideriksen et al., 2023). Consistently and across languages, entrainment rates are higher in affiliative conversations, while entrainment levels are higher in task-oriented conversations. Furthermore, we see differences in entrainment levels between the Map Task and the Alien Game. The Alien Game drives the high occurrence of lexical entrainment level, while the numbers for the Map Task are lower than for affiliative conversations. Once again, the asymmetry in access to information between the two tasks seems to afford different strategies for building mutual understanding. This further supports the argument that conversational devices are modulated by context; in particular, task-oriented conversations show less frequent entrainment (lower rate), but when there is entrainment, it is more substantial (higher level), thus—speculatively—meeting the increased demands for referential specificity.

Interestingly, we see similar effects for entrainment rate and level (both higher in Danish than Norwegian), where previous studies show important differences in their behavior related to the context (Dideriksen et al., 2023) or the development of linguistic and conversational skills (Fusaroli et al., 2023a, 2023b; Misiek, Favre, & Fourtassi, 2020). However, as in previous literature, rate presents bigger variations related to language and context than level does. This could indicate that both rate and level can provide the redundancy needed to cope with sound opacity (and possibly other communication issues), but they do so in different ways.

Summing up, native speakers of Danish entrain more than Norwegians, possibly for the purpose of creating an entrenched, more redundant context to facilitate the processing and interpretation of new utterances. Given that linguistic forms are often repeated, the speech signal is less challenging to process. This might explain the unexpected results for repairs: continuous entrainment might obviate the need for repairs, which are indeed less frequent in Danish than in Norwegian. A post hoc analysis of the relation between the use of repair and different measures of entrainment reveals small but credible and consistent negative correlation scores between -0.18 and -0.26 (ERs > 20) between the two devices. In other words, interlocutors and pairs using more entrainment within a specific conversation tend also to use fewer repairs within that conversation. This suggests that cross-linguistic differences in the use of entrainment might explain some of the cross-linguistic differences in the use of repair (although not all of them as indicated by the small effect sizes of the correlations).

4.1.4. Relations between conversational devices

As the three conversational devices investigated are likely to be interrelated in their use (Fusaroli et al., 2017), we performed exploratory analyses of such relationships. Considering backchannels and repairs, we only see four cases in the full corpus of turns being double-coded as backchannel and repair, two in Danish and two in Norwegian, which seems negligible. At the individual level, the use of repairs was only weakly positively associated with the use of backchannels (<1% of the variance explained). Regarding backchannels relation to entrainment, we observe at the turn level that backchannels involve lower entrainment of any form, compared to non-backchannel turns; and the same is true for open repairs and entrainment. Restricted requests and restricted offers were associated with higher lexical (but neither syntactic nor semantic) entrainment than other turns, especially in task-oriented conversations, across languages. However, conversation-level analyses indicated that the general frequency of repair (and not of specific forms of repair) was negatively associated with all forms of entrainment. Yet, none of these turn- or conversation-level associations explained more than 4% of the variance in the data, thus suggesting that such overlapping only has a minor explanatory role.

4.2. The interplay of conversational, linguistic, and cognitive dynamics

Following up on the notion of language as a complex adaptive system (Beckner et al., 2009), we suggest that we are seeing a complex interaction of conversational, linguistic, and cognitive dynamics at work. Previous work has suggested that native speakers of Danish compensate for the relative acoustic opacity of their language by relying more on top-down contextual information (Ishkhanyan et al., 2020; Trecca et al., 2021). The current theoretical proposal and results provide a complementary perspective to that: to compensate for the particular features of their language, native speakers of Danish seem to build more redundant contextual niches in which words, structures, and topics are constantly revisited. This strategy is adjusted to the needs of the ongoing activities (here, task-oriented conversations vs. affiliative conversations) but always within the longer-term constraints deriving from language-specific features such as the relative acoustic opacity of the language. For instance, the Danish-speaking participants did not decrease entrainment during task-oriented conversations to the extent Norwegian speakers did. These findings thus motivate further elaborations of previous theoretical frameworks for investigating conversational devices (Dideriksen et al., 2023), adding a layer of global constraints pertaining to the impact of language-specific features.

To this end, we here advance the idea of interactions and feedback loops between the phenomena at play (see Fig. 9). Language-specific properties (here, sound opacity) create a linguistic environment that favors context-driven inferences and increased linguistic entrainment. Context-driven inference is facilitated by linguistic entrainment, which creates a more informationally redundant niche for easier disambiguation of the speech signal, and potential identification of misunderstandings (Bjørndahl, Fusaroli, Østergaard, & Tylén, 2015). Thus, cognitive, and conversational strategies might reinforce each other. Further, the construction of an increased reliance on a richer context might reduce some of the pressure against the

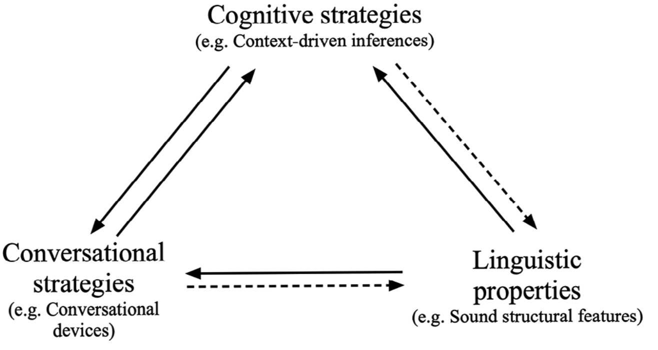

Fig. 9. A diagram of the interplay between cognitive strategies, conversational strategies, and linguistic properties. While the components are presented as separate units, we fully acknowledge that they are not completely separable, and that, for instance, cognitive and conversational strategies are deeply intertwined. The arrows indicate the direction of influence, with bold arrows indicating a more direct influence (on a short time scale), while dashed arrows depict a more indirect influence (on a longer time scale).

persistence of Danish sound opacity by diminishing its impact on successful communication. Thus, we hypothesize a feedback loop between opacity and compensatory strategies, the former selecting for compensation and the latter enabling albeit on a different time scale, the persistence of opacity.

By developing this framework and refining it according to the results of the investigation, we provide both theoretical and empirical contributions to the debate. We acknowledge that the connections between sound opacity and specific patterns of conversational dynamics are still tentative. However, by making our hypotheses explicit, assessing them, discarding some (e.g., sound opacity leading to more frequent repairs and backchannels) and refining others, we provide valuable theory development and more specific theories, which can be challenged. We recommend for the findings to be replicated and for their generalizability to be assessed: Does the pattern hold for other sound-opaque languages? Does the pattern hold when artificially increasing sound opacity—that is, at short time scales? Can alternative explanations (e.g., related to other cultural factors) be ruled out? If we conceive of scientific research as a collective and cumulative enterprise, where no single study (or even sequence of studies) can provide definitive evidence, our study does provide tentative theory development, and with that, additional steps, and concrete pointers for further challenges to and developments of the theory.

5. Limitations

There are several limitations to keep in mind when interpreting the results of this study. First, we assumed minimal differences between Norwegian and Danish, besides phonetic reduction. However, additional differences might be present, which we have not accounted for. For example, Norway is known for its wide variety of dialects across the country

(Kristoffersen, 2000). This contrasts with Danish, which has comparatively little dialectal variance across regional areas (Basbøll, 2005). The Norwegian data collection in the current study was carried out in the Bergen area of Western Norway, but participants were students coming from different regions in Norway with differing dialectal variations. Such dialectal variation in the corpus could have affected the outcome of our study. Further explorations of the corpus should control for dialectal variance to assess the effect of dialect on the establishment of mutual understanding across conversational contexts. Additionally, we also observed non-trivial differences in speaker talkativeness between the two languages. While the number of utterances was comparable, native speakers of Danish produced about 50% more conversational turns, which were also two words longer on average, thus producing 75% more words than Norwegians for a comparable number of conversations and conversational duration. While differences in utterance length do not affect our results—due to our within-language surrogate baselines—they do suggest additional variation in conversational practice that we had not anticipated and warrant further investigation. Along the same lines, it is important to note that the above analysis cannot make any assumptions about qualitative cultural differences in the functional use of conversational devices between Danish and Norwegian. Backchannels have, for instance, been found to have varying functions in conversations cross-culturally (Cutrone, 2005, 2011; Li et al., 2010; Maynard, 1990; White, 1989), and an obvious next step would be to conduct a more qualitative analysis of the function of conversational devices across languages.

Second, our analysis focused exclusively on *verbal* devices and how they adapt to contextual and linguistic constraints on a device-by-device basis. We do not consider how, for instance, more specific forms of repair often involve linguistic entrainment and how that might impact our estimates, compared to an analysis of entrainment where repairs are removed. Further, an increasing body of research has underscored the importance of multimodal mechanisms such as prosody, turn-taking dynamics, head nods, hand gestures, and eye blinks in conversation (Holler & Levinson, 2019; Louwerse et al., 2012). It is very likely that these non-verbal mechanisms are also at play, for instance, in the form of gestural entrainment or non-verbal backchannels and repair initiations. Further, there is increasing evidence of the importance of prosodic and acoustic entrainment, its cross-linguistic and contextual variability and its relations to performance in task-oriented conversations (Fusaroli & Tylén, 2015; Levitan et al., 2015; Ostrand & Chodroff, 2021; Savino, Lapertosa, Caffò, & Refice, 2016; Wynn, Barrett, Berisha, Liss, & Borrie, 2023). Investigating multimodal aspects of interaction will add crucial knowledge about the complexity behind establishing and maintaining mutual understanding in a conversation (e.g., Trujillo, Dideriksen, Tylén, Christiansen, & Fusaroli, 2023).

Third, the distinction between task-oriented and affiliative conversations might not capture the multiplicity of motives in a conversation. We might expect that instead of a dichotomy, conversational goals occur on a continuous scale, where the purpose of the conversation can be multifaceted within an affiliative or task-oriented conversation (cf. the conversational circumplex in Yeomans, Schweitzer, & Brooks, 2022). Thus, a more graded definition of both task and affiliative conversations is needed to fully capture the underlying diversity of the use of conversational devices. This could be tested by systematically manipulating the contextual constraints on a continuous scale, where the amount of shared visual references within the same task differs. This would make it possible to make stronger inferences about the relationship between contextual demands and the use of conversational devices.

Fourth, our interpretation of the results should be further explored by a more direct quantification of consonantal reduction and sound opacity as well as by developing experimental manipulations of the causal elements. Related to the former, one would need to develop and validate automated measures of consonantal reduction and assess the impact of different levels of reduction on the perception of isolated sounds and words as well as the same words in larger contexts (e.g., conversational ones), as well as how it relates to conversational dynamics (for an initial exploration, see Ishkhanyan et al., 2020). Additional experimental manipulations could involve, for instance, artificially increasing certain forms of sound opacity via background noise or speech filtering, in which case, we would expect the use of conversational devices to change to compensate for the ambiguous input. Until recently, manipulations such as these would mostly be applied within single-listener paradigms (Trecca et al., 2021; Yurovsky, Case, & Frank, 2017). However, a recent study applied varying degrees of noise in conversations between dyads and triads and found increased coherence of bodily movement in conversations with high levels of background noise (Hadley & Ward, 2021). More studies are needed to investigate if these effects differ cross-linguistically and whether slow adaptation (due to linguistic structures) mirrors fast adaptation (due to local experimental manipulations).

Finally, more work is needed to replicate our findings and, more crucially, extend them to a broader range of linguistic and cultural elements to assess their generalizability. Other languages might present similar or comparable features in their sound structure, for instance, phonetic reduction of consonants in Mandarin or vowel elision in European Portuguese (compared to Brazilian Portuguese), and therefore be amenable to theory-motivated cross-linguistic contrasts of conversational devices (see also Christiansen et al., 2022; Deffner et al., 2021). Broader cross-cultural comparisons in the use of conversational devices could also benefit from a theory-driven perspective identifying relevant dimensions of variation (e.g., see Muthukrishna et al., 2020, for an example).

6. Implications

Our findings highlight not only how local task-related and more global linguistic constraints affect conversational dynamics but also potential trade-offs between the two. For instance, while a high diversity of different contributions would seem to be a good strategy in task-oriented conversations, it should be tempered when mutual understanding might be at risk, for instance, when the speech signal is more challenging to process (see also Hadley & Ward, 2021, for a similar finding on head movement synchrony). Highlighting the need for careful consideration of trade-offs between different constraints has implications for a range of fields where in-depth knowledge about conversational patterns is crucial:

1. The increased use of interactive virtual assistants in home environments requires new and more advanced standards within the field of human–computer interaction. Knowledge about conversational patterns can help develop more realistic, useful, and

advanced dialogue systems, with more complex turn-taking structures that are better adapted to contextual and linguistic constraints (Loth et al., 2015; Ruane, Birhane, & Ventresque, 2019; Sugiyama et al., 2018; Yan, 2018). For instance, future developments in human–computer interaction could include repair sequences to reduce the number of unsuccessful interactions with voice assistants and varying levels/rates of linguistic entrainment in virtual dialogue systems adjusted to the context in which the conversation is happening.

2. Neuropsychiatric conditions, such as autism and schizophrenia, are often construed as a form of social atypicality; however, we know very little about how this actually unfolds in social interactions (Christiansen & Chater, 2022; Dwyer et al., 2019; Fusaroli et al., 2017, 2019; Hopkins et al., 2016). We speculate that conceptualizing conversations as complex adaptive systems might provide a more inclusive perspective and ultimately a deeper understanding of such atypicality, and the interactional failures and successes involved. Neurodiversity might pose diverse constraints on conversations, including mismatched expectations (Milton, 2012; Milton, Gurbuz, & López, 2022), which might lead to diverse and perhaps mismatched use of conversational strategies across both interlocutors (Di Paolo, Cuffari, & De Jaegher, 2018). By developing a more nuanced and systematic understanding of conversational patterns across diverse constraints, we can provide a much better insight into how these patterns might differ in conversations involving neurodiverse people and the potential impact of this diversity.
3. More generally, the current study emphasizes the importance of cross-linguistic perspectives on the study of conversations, either within one study or cumulatively by more explicitly building upon previous studies and comparing the findings. We speculate that these changes in conversational strategies might go hand in hand with changes in cognitive strategies (Trecca et al., 2021), but more research is needed.

7. Conclusion

To our knowledge, these findings are the first to experimentally manipulate both local (task vs. affiliative) and global (Danish vs. Norwegian) dialogue contexts to reveal conversational differences in strategies that lead to mutual understanding. By adopting a rigorous framework for investigating conversational devices in conversations, we were able to detect both short- and long-term likely constraints on the establishment of mutual understanding in interactions across Danish and Norwegian. The finding that sound opacities appear to lead to language-specific adaptive modulation of these processes provides a strong empirical foundation for future research to delve further into the systematic variations in language use caused by external constraints.

Acknowledgments

This study was supported by the Danish Council for Independent Research (FKK) Grant DFF- 7013-00074 awarded to Morten H. Christiansen and seed funding grants from the

Interacting Minds Centre at Aarhus University. Mark Dingemanse acknowledges funding from NWO grant 016.vidi.185.205. We thank Patrick Healey and Sophie Skach for providing expert contributions to the informed priors; Jakob Steensig and Kathrine Garly for defining and discussing coding schemes; and the many research assistants involved in the transcribing and coding of the datasets (see Section S4 in the Supporting Information for a full list of names).

Open Research Badges

This article has earned Open Data badge. Data is available at https://osf.io/x3s6w/.

Data Availability Statement

This study was not preregistered. All data have been made publicly available at Open Science Framework and can be accessed at: https://osf.io/x3s6w/.

Supplemental materials

https://osf.io/x3s6w/

References

Albert, S., & de Ruiter, J. (2018). Improving human interaction research through ecological grounding. *Collabra: Psychology*, *4*(1), 24.

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., & Miller, J. (1991). The HCRC map task corpus. *Language and Speech*, *34*(4), 351–366.

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, *27*(2), 195–225.

Basbøll, H. (2005). *The phonology of Danish*. Oxford, England: Oxford University Press.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., & Larsen-Freeman, D. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*(s1), 1–26.

Benus, S., Gravano, A., & Hirschberg, J. B. (2007). *The prosody of backchannels in American English*. https://academiccommons.columbia.edu/doi/10.7916/D8BK1MMW/download and it's been accessed on nov 20 2023

Bertrand, R., & Espesser, R. (2017). Co-narration in French conversation storytelling: A quantitative insight. *Journal of Pragmatics*, *111*, 33–53.

Bjørndahl, J., Fusaroli, R., Østergaard, S., & Tylén, K. (2015). When agreeing is not enough. *Interaction Studies*, *16*(3), 495–525.

Bleses, D., Vach, W., Slott, M., Wehberg, S., Thomsen, P. I. A., Madsen, T. O., & Basbøll, H. (2008). Early vocabulary development in Danish and other languages: A CDI-based comparison. *Journal of Child Language*, *35*(3), 619–650.

Bleses, D., & Basbøll, H. (2004). The Danish sound structure: Implications for language acquisition in normal and hearing impaired populations. *In Brain, Hearing and Learning. Proceedings of the 20th Danavox Symposium* (pp. 165–190).

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32. https://doi.org/10.18637/jss.v076.i01

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Chartrand, T. L., & Lakin, J. L. (2013). The antecedents and consequences of human behavioral mimicry. *Annual Review of Psychology*, *64*, 285–308.

Christiansen, M. H., & Chater, N. (2016). *Creating language: Integrating evolution, acquisition, and processing*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2022). *The language game: How improvisation created language and changed the world*. New York: Random House.

Christiansen, M. H., Contreras Kallens, P., & Trecca, F. (2022). Toward a comparative approach to language acquisition. *Current Directions in Psychological Science*, *31*(2), 131–138. https://doi.org/10.1177/09637214211049229

Clark, H. H. (2009). Context and common ground. In J. L. Mey (Ed.), *Concise encyclopedia of pragmatics* (pp. 116–119). Amsterdam: Elsevier.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. *Perspectives on Socially Shared Cognition*, *13*(1991), 127–149.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259–294.

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

Colman, M., & Healey, P. (2011). The distribution of repair in dialogue. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, Austin, TX.

Cox, C., Bergmann, C., Fowler, E., Keren-Portnoy, T., Roepstorff, A., Bryant, G. A., & Fusaroli, R. (2022a). A systematic review and Bayesian meta-analysis of the acoustic features of infant-directed speech. *Nature Human Behavior*. *7*, 114–133. https://doi.org/10.31234/osf.io/vn62t

Cox, C., Dideriksen, C., Keren-Portnoy, T., Roepstorff, A., Christiansen, M. H., & Fusaroli, R. (2022b). Infant-directed speech does not always involve exaggerated vowel distinctions: Evidence from Danish. *Child Development*. https://doi.org/10.1111/cdev.13950

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, *1695*(5), 1–9.

Cutrone, P. (2005). A case study examining backchannels in conversations between Japanese–British dyads. *Multilingua—Journal of Cross-Cultural and Interlanguage Communication*, *24*(3), 237–274.

Cutrone, P. (2011). Politeness and face theory: Implications for the backchannel style of Japanese L1/L2 Speakers. *Language Studies Working Papers*, *3*, 51–57.

Dale, R., Fusaroli, R., Duran, N. D., & Richardson, D. C. (2013). The self-organization of human interaction. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. *59*, pp. 43–95). Amsterdam: Elsevier.

Dale, R., Fusaroli, R., Håkonsson, D. D., Healey, P., Mønster, D., McGraw, J. J., Mitkidis, P., & Tylén, K. (2013). Beyond synchrony: Complementarity and asynchrony in joint action. *Proceedings of the Cognitive Science Conference*, Berlin, Germany.

Deffner, D., Rohrer, J. M., & McElreath, R. (2021). A causal framework for cross-cultural generalizability. https://doi.org/10.1177/25152459221106366

Di Paolo, E. A., Cuffari, E. C., & De Jaegher, H. (2018). *Linguistic bodies: The continuity between life and language*. Cambridge, MA: MIT Press.

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), 864–889. https://doi.org/10.1037/xge0001301

Dideriksen, C., Fusaroli, R., Tylén, K., Dingemanse, M., & Christiansen, M. H. (2019). Contextualizing conversational strategies: Backchannel, repair and linguistic alignment in spontaneous and task-oriented conversations. *CogSci' 19*, Montreal, Canada (pp. 261–267).

Dingemanse, M., Roberts, S. G., Baranova, J., Blythe, J., Drew, P., Floyd, S., Gisladottir, R. S., Kendrick, K. H., Levinson, S. C., & Manrique, E. (2015). Universal principles in the repair of communication problems. *PLoS One*, *10*(9), e0136100.

Drummond, K., & Hopper, R. (1993). Back channels revisited: Acknowledgment tokens and speakership incipiency. *Research on Language and Social Interaction*, *26*(2), 157–177.

Duran, N. D., Paxton, A., & Fusaroli, R. (2019). ALIGN: Analyzing linguistic interactions with generalizable techNiques—A Python library. *Psychological Methods*, *4*(4), 419–438.

Dwyer, K., David, A. S., McCarthy, R., McKenna, P., & Peters, E. (2019). Linguistic alignment and theory of mind impairments in schizophrenia patients' dialogic interactions. *Psychological Medicine*, *50*(13), 2194–2202.

Fay, N., Ellison, T. M., Tylén, K., Fusaroli, R., Walker, B., & Garrod, S. (2018). Applying the cultural ratchet to a social artefact: The cumulative cultural evolution of a language game. *Evolution and Human Behavior*, *39*(3), 300–309.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931–939.

Fusaroli, R., Bjørndahl, J. S., Roepstorff, A., & Tylén, K. (2016). A heart for interaction: Shared physiological dynamics and behavioral coordination in a collective, creative construction task. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1297–1310.

Fusaroli, R., Demuru, P., & Borghi, A. M. (2012). The intersubjectivity of embodiment. *Journal of Cognitive Semiotics*, *4*(1), 1–5.

Fusaroli, R., Gangopadhyay, N., & Tylén, K. (2014). The dialogically extended mind: Making a case for language as skilful intersubjective engagement. *Cognitive Systems Research*, *29-30*, 31–39.

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*, 147–157.

Fusaroli, R., & Tylén, K. (2012). Carving language for social coordination: A dynamic approach. *Interaction Studies*, *13*, 103–123.

Fusaroli, R., & Tylén, K. (2015). Investigating conversational dynamics: Interactive alignment, interpersonal synergy, and collective task performance. *Cognitive Science*, *40*(1), 145–171. https://doi.org/10.1111/cogs.12251

Fusaroli, R., Tylén, K., Garly, K., Steensig, J., Christiansen, M. H., & Dingemanse, M. (2017). Measures and mechanisms of common ground: Backchannels, conversational repair, and interactive alignment in free and task-oriented social interactions. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th annual conference of the cognitive science society (CogSci 2017)* (pp. 2055–2060). Austin, TX: Cognitive Science Society.

Fusaroli, R., Weed, E., Fein, D., & Naigles, L. (2019). Hearing me hearing you: Reciprocal effects between child and parent language in autism and typical development. *Cognition*, *183*, 1–18.

Fusaroli, R., Weed, E., Rocca, R., Fein, D., & Naigles, L. (2023a). Caregiver linguistic alignment to autistic and typically developing children. *Cognition*, *236*, 105422.

Fusaroli, R., Weed, E., Rocca, R., Fein, D., & Naigles, L. (2023b). Repeat after me? Both children with and without autism commonly align their language with that of their caregiver. *Cognitive Science*, *47*(11), e13369. https://doi.org/10.31234/osf.io/m8fhk

Gardner, R. (2001). *When listeners talk*. Amsterdam: Benjamins.

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences*, *8*(1), 8–11.

Gooskens, C. (2006). Linguistic and extra-linguistic predictors of inter-Scandinavian intelligibility. *Linguistics in the Netherlands*, *23*(1)101–113.

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, *34*(4), 365–399.

Grønnum, N. (1998). Danish. *Journal of the International Phonetic Association*, *28*(1/2), 99–105.

Grønnum, N. (2005). *Fonetik og fonologi: Almen og dansk*. Copenhagen, Denmark: Akademisk.

Hadley, L. V., & Ward, J. A. (2021). Synchrony as a measure of conversation difficulty: Movement coherence increases with background noise level and complexity in dyads and triads. *PLoS One*, *16*(10), e0258247.

Healey, P. G., Mills, G. J., Eshghi, A., & Howes, C. (2018). Running repairs: Coordinating meaning in dialogue. *Topics in Cognitive Science*, *10*(2), 367–388.

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

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 61–83.

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, *23*(8), 639–652.

Hömke, P., Holler, J., & Levinson, S. C. (2018). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS One*, *13*(12), e0208030.

Hopkins, Z., Yuill, N., & Keller, B. (2016). Children with autism align syntax in natural conversation. *Applied Psycholinguistics*, *37*(2), 347–370.

Howes, C., & Eshghi, A. (2021). Feedback relevance spaces: Interactional constraints on processing contexts in Dynamic Syntax. *Journal of Logic, Language and Information*, *30*(2), 331–362.

Ireland, M. E., & Henderson, M. D. (2014). Language style matching, engagement, and impasse in negotiations. *Negotiation and Conflict Management Research*, *7*(1), 1–16.

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J., & Pennebaker, J. W. (2011). Language style matching predicts relationship initiation and stability. *Psychological Science*, *22*(1), 39–44.

Ishkhanyan, B., Fusaroli, R., Højen, A., Tylén, K., Johansson, C., & Christiansen, M. H. (2020). *Cross-linguistic differences in categorical perception: Comparison of Danish and Norwegian*.

Kirkedal, A., Plank, B., Derczynski, L., & Schluter, N. (2019). The lacunae of Danish natural language processing. *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, Turku, Finland (pp. 356–362).

Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford, England: Oxford University Press on Demand.

Levitan, R., Beňuš, Š., Gravano, A., & Hirschberg, J. (2015). Acoustic-prosodic entrainment in Slovak, Spanish, English and Chinese: A cross-linguistic comparison. *Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, Prague, Czech Republic (pp. 325–334).

Li, H. Z., Cui, Y., & Wang, Z. (2010). Backchannel responses and enjoyment of the conversation: The more does not necessarily mean the better. *International Journal of Psychological Studies*, *2*(1), 25.

Loth, S., Jettka, K., Giuliani, M., & De Ruiter, J. P. (2015). Ghost-in-the-machine reveals human social signals for human–robot interaction. *Frontiers in Psychology*, *6*, 1641.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426.

Mattys, S. L., & Jusczyk, P. W. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance*, *27*(3), 644–655.

Maynard, S. K. (1990). Conversation management in contrast: Listener response in Japanese and American English. *Journal of Pragmatics*, *14*(3), 397–412.

Micklos, A., & Woensdregt, M. (2023). Cognitive and interactive mechanisms for mutual understanding in conversation. In G. W. Nobit (Ed.), *Oxford research encyclopedia of communication*. Oxford, England: Oxford University Press.

Mills, G., Groningen, C., & Redeker, G. (2017). Amplifying signals of misunderstanding improves coordination in dialogue. *FADLI 2017*, Toulouse, France (pp. 52–54).

Milton, D. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society*, *27*(6), 883–887.

Milton, D., Gurbuz, E., & López, B. (2022). The 'double empathy problem': Ten years on. *Autism*, *26*(8), 1901–1903.

Misiek, T., Favre, B., & Fourtassi, A. (2020). Development of multi-level linguistic alignment in child-adult conversations. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, Online (pp. 54–58).

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., & Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance. *Psychological Science*, *31*(6), 678–701.

Nazzi, T., Dilley, L. C., Jusczyk, A. M., Shattuck-Hufnagel, S., & Jusczyk, P. W. (2005). English-learning infants' segmentation of verbs from fluent speech. *Language and Speech*, *48*(3), 279–298.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360.

Ostrand, R., & Chodroff, E. (2021). It's alignment all the way down, but not all the way up: Speakers align on some features but not others within a dialogue. *Journal of Phonetics*, *88*, 101074. https://doi.org/10.1016/j.wocn.2021.101074

Peirce, J., & MacAskill, M. (2018). *Building experiments in PsychoPy*. Los Angeles, CA: Sage.

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

Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge, MA: Cambridge University Press.

Purver, M., Ginzburg, J., & Healey, P. (2003). On the Means for Clarification in Dialogue. In J. van Kuppevelt & R. W. Smith (Eds.), *Current and new directions in discourse and dialogue* (pp. 235–255). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-010-0019-2_11

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

Reitter, D., & Moore, J. D. (2006). Priming of syntactic rules in task-oriented dialogue and spontaneous conversation. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Vancouver, Canada.

Reitter, D. & Moore, J. D. (2014). Alignment and task success in spoken dialogue. *Journal of Memory and Language*, *76*, 29–46.

Ruane, E., Birhane, A., & Ventresque, A. (2019). Conversational AI: Social and Ethical Considerations. *27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science*, Galway, Ireland (pp. 104–115).

Savino, M., Lapertosa, L., Caffò, A., & Refice, M. (2016). Measuring prosodic entrainment in Italian collaborative game-based dialogues. *Speech and Computer: 18th International Conference, SPECOM 2016*, Budapest, Hungary (pp. 476–483).

Scheflen, A. E. (1964). The significance of posture in communication systems. *Psychiatry*, *27*(4), 316–331.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh'and other things that come between sentences. In D. Tannen (Ed.), *Analyzing discourse: Text and talk* (pp. 71–93). Washington, DC: Georgetown University Press.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414.

Shockley, K., Santana, M.-V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326–332.

Slocombe, K. E., Alvarez, I., Branigan, H. P., Jellema, T., Burnett, H. G., Fischer, A., Li, Y. H., Garrod, S., & Levita, L. (2013). Linguistic alignment in adults with and without Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *43*(6), 1423–1436.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., De Ruiter, J. P., & Yoon, K.-E. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587–10592.

Straka, M., Hajic, J., & Straková, J. (2016). UDPipe: Trainable pipeline for processing CoNLL-U files performing tokenization, morphological analysis, pos tagging and parsing. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, Portorož, Slovenia (pp. 4290–4297).

Strømberg-Derczynski, L. (2018). Dapipe. GitHub Repository. Available at: https://Github.Com/ITUnlp/Dapipe. accessed on nov 20 2023

Strømberg-Derczynski, L., Baglini, R., Christiansen, M. H., Ciosici, M. R., Dalsgaard, J. A., Fusaroli, R., Henrichsen, P. J., Hvingelby, R., Kirkedal, A., & Kjeldsen, A. S. (2020). The Danish Gigaword Project. Available at: arXiv Preprint arXiv:2005.03521. accessed on nov 20 2023

Sugiyama, H., Meguro, T., Yoshikawa, Y., & Yamato, J. (2018). *Improving dialogue continuity using inter-robot interaction, 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, Nanjing, China (pp. 105–112).

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.

Trecca, F., Bleses, D., Madsen, T. O., & Christiansen, M. H. (2018). Does sound structure affect word learning? An eye-tracking study of Danish learning toddlers. *Journal of Experimental Child Psychology*, *167*, 180–203.

Trecca, F., Tylén, K., Højen, A., & Christiansen, M. H. (2021). Danish as window onto language processing and learning. *Language Learning*, *71*(3), 799–833.

Trujillo, J., Dideriksen, C., Tylén, K., Christiansen, M. H., & Fusaroli, R. (2023). The dynamic interplay of kinetic and linguistic coordination in Danish and Norwegian conversation. *Cognitive Science*, *47*(6), e13298.

Truong, K. P., & Heylen, D. (2010). Disambiguating the functions of conversational sounds with prosody: The case of 'yeah'. *Eleventh Annual Conference of the International Speech Communication Association*, Makuhari, Chiba, Japan.

Tylén, K., Fusaroli, R., Bundgaard, P., & Østergaard, S. (2013). Making sense together: A dynamical account of linguistic meaning making. *Semiotica*, *194*(1/4), 39–62.

Tylén, K., Fusaroli, R., Østergaard, S. M., Smith, P., & Arnoldi, J. (2023). The social route to abstraction: Interaction and diversity enhance performance and transfer in a rule-based categorization task. *Cognitive Science*, *47*(9), e13338. https://doi.org/10.1111/cogs.13338

White, S. (1989). Backchannels across cultures: A study of Americans and Japanese. *Language in Society*, *18*(1), 59–76.

Wynn, C. J., Barrett, T. S., Berisha, V., Liss, J. M., & Borrie, S. A. (2023). Speech entrainment in adolescent conversations: A developmental perspective. *Journal of Speech, Language, and Hearing Research*, *66*(8S), 3132–3150. https://doi.org/10.1044/2023_JSLHR-22-00263

Wynn, C. J., & Borrie, S. A. (2022). Classifying conversational entrainment of speech behavior: An expanded framework and review. *Journal of Phonetics*, *94*, 101173. https://doi.org/10.1016/j.wocn.2022.101173

Yan, R. (2018). " Chitty-chitty-chat bot": Deep learning for conversational AI. *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI, 2018)*, Stockholm, Sweden (pp. 5520–5526).

Yeomans, M., Schweitzer, M. E., & Brooks, A. W. (2022). The Conversational Circumplex: Identifying, prioritizing, and pursuing informational and relational motives in conversation. *Current Opinion in Psychology*, *44*, 293–302.

Yngve, V. H. (1970). On getting a word in edgewise. *Sixth Regional Meeting of the Chicago Linguistics Society*, Chicago, IL (pp. 567–577).

Yurovsky, D., Case, S., & Frank, M. C. (2017). Preschoolers flexibly adapt to linguistic input in a noisy channel. *Psychological Science*, *28*(1), 132–140.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information