ASHA American Speech-Language-Hearing Association

JSLHR

Research Note

Studying Conversational Adjustments in Interaction: Beyond Acoustic Phonetic Changes

Catherine T. Pham[a] and Navin Viswanathan[b]

[a] Department of Psychology, The Pennsylvania State University, University Park [b] Department of Communication Sciences and Disorders, The Pennsylvania State University, University Park

ARTICLE INFO

Article History:
Received April 21, 2023
Revision received August 25, 2023
Accepted October 1, 2023

Editor-in-Chief: Julie A. Washington
Editor: Reem Khamis

https://doi.org/10.1044/2023_JSLHR-23-00268

ABSTRACT

Purpose: We examined which measures of complexity are most informative when studying language produced in interaction. Specifically, using these measures, we explored whether native and nonnative speakers modified the higher level properties of their production beyond the acoustic–phonetic level based on the language background of their conversation partner.
Method: Using a subset of production data from the Wildcat Corpus that used Diapix, an interactive picture matching task, to elicit production, we compared English language production at the dyad and individual level across three different pair types: eight native pairs (English–English), eight mixed pairs (four English–Chinese and four English–Korean), and eight nonnative pairs (four Chinese–Chinese and four Korean–Korean).
Results: At both the dyad and individual levels, native speakers produced longer and more clausally dense speech. They also produced fewer silent pauses and fewer linguistic mazes relative to nonnative speakers. Speakers did not modify their production based on the language background of their interlocutor.
Conclusions: The current study examines higher level properties of language production in true interaction. Our results suggest that speakers' productions were determined by their own language background and were independent of that of their interlocutor. Furthermore, these demonstrated promise for capturing syntactic characteristics of language produced in true dialogue.
Supplemental Material: https://doi.org/10.23641/asha.24712956

Typical language use often involves two (or more) people interacting with one another, shifting between the role of speaker and listener. These interactions are complex and dynamic, involving a range of communicative behavior like turn-taking, negotiations of meaning, and the establishment of shared knowledge and context. Effective communication has been argued to depend on the coordination and/or synchronization of speakers' linguistic behaviors (Fusaroli et al., 2014; Pickering & Garrod, 2004). As such, understanding how interlocutors coordinate with each other by adjusting their production in interaction is crucial for understanding how individuals effectively communicate.

Despite the fact that language use is typically interactive, spoken language research at the phonological, syntactic, and even pragmatic levels has largely focused on studying the behaviors of individual language users. This approach to the study of language use, referred to as the language-as-product approach (see Clark, 1992, for further discussion), relies on highly specific language tasks (often in the absence of a communicative goal) to investigate the cognitive processes underlying language abilities. Although the product approach allows for precise control over experimental conditions and stimuli, the extent to which language use in monologue generalizes to dialogue—a more naturalistic communicative setting—is unclear. In contrast, the language-as-action approach focuses on interactive processes by considering how the demands of the social and physical environment shape language use (Clark, 1992). Unlike the product approach, which utilizes scripted utterances to probe the language user's comprehension and production, the action approach examines comprehension and production by examining how pairs of interlocutors collaborate during a communicative task.

Correspondence to Catherine T. Pham: catherine.pham@psu.edu.
Disclosure: *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

One type of experimental paradigm that allows researchers to study how individuals use language in collaborative contexts is the use of goal-oriented tasks to elicit dialogue (e.g., Diapix task, map task, and tangram task). These tasks allow for a fair level of control over the semantic content of speakers' productions, allowing researchers to target the production of specific words and/or sounds. As such, many studies have successfully utilized these experimental paradigms to focus on examining production changes at the acoustic–phonetic level (e.g., Hazan & Baker, 2011; Iyer et al., 2016; Lee & Baese-Berk, 2020; Pardo, 2006; Tuomainen et al., 2021; Van Engen et al., 2010). Although these studies provide clear evidence that speakers acoustically modify their productions for the benefit of their addressee, much less work has examined higher level modifications in interaction that go beyond the acoustic–phonetic level.

Notably, one prominent task that has gained popularity in recent years is the Diapix task (Baker & Hazan, 2011; Van Engen et al., 2010), which requires pairs of participants work together to identify differences between their two images. Although this task was originally designed to examine production at the acoustic–phonetic level, we believe that the Diapix task can be extended to the study of syntactic production as well. As highlighted by Van Engen et al. (2010), the Diapix task provides several benefits over other task-based dialogue elicitation techniques (e.g., map task; Anderson et al., 1991). The Diapix task allows for more balanced contributions between the two participants, and, of most interest to the current study, the Diapix task elicits a range of utterance types such as declarative statements, questions and answers, imperatives (e.g., "Tell me what you see in your picture"), and exclamations/interjections (e.g., "ah ha," "great," and "oops"). The variety of utterances elicited would presumably lend itself to the study of speaker modifications in syntax. However, to our knowledge, interactive tasks, in general, have not yet been utilized to examine syntactic adjustments in dialogue. In this study, we focus on characterizing conversational adjustments in syntactic complexity and related changes. We concentrate on these analyses as a complement to existing acoustic–phonetic analyses of conversational adjustments, because syntax is one of the most malleable properties of language output. Speakers can, and readily do, dramatically alter the way they combine words and phrases (e.g., MacDonald, 2013). Thus, for a complete understanding of speaker adjustments, we need a reliable way of assessing speech changes in actual interaction at this level.

Defining and Operationalizing Syntactic Complexity

The notion of syntactic complexity has been extensively discussed across various disciplines within the language sciences (e.g., formal theoretical linguistics, comparative linguistics, and language typology). However, we focus on how the fields of second language (L2) acquisition (SLA)—and relatedly first-language acquisition and speech, language, and hearing science (SLHS)—have approached the study of complexity. We primarily concentrate on work conducted in these disciplines, because they have generally adopted a more applied approach to the study of complexity by focusing on developing complexity measures to capture language users' performance, proficiency, and development and evaluating how language ability may change over time or across different communicative contexts (see Housen et al., 2019).

A distinction has been drawn between absolute complexity and relative complexity (Miestamo, 2008; see also Housen et al., 2019). Absolute complexity refers to the elaborateness and variety of the components within a linguistic construction or system (see Miestamo, 2009). Regarding measures of absolute complexity, commonly adopted metrics include length-based measures (e.g., mean length of sentence/phrase/clause), with longer lengths reflecting greater complexity, and measures capturing the degree of embeddedness and/or coordination of linguistic constituents into larger constituents (e.g., number of dependent clauses/phrases per sentence), with greater number of clauses/phrases reflecting greater complexity (for reviews of measures of [absolute] complexity, see Jagaiah et al., 2020; Ortega, 2003; Wolfe-Quintero, 1998).

In contrast, relative complexity refers to the amount of cognitive demand placed upon the language user. Thus, a linguistic feature is considered complex if it is cognitively taxing to learn, process, and/or produce (Bulté & Housen, 2012). Within first (and to some extent second) language acquisition research, relative complexity has been operationalized using various indices of language development (e.g., Developmental Sentence Scoring: Lee & Canter, 1971; Index of Productive Syntax: Scarborough, 1990, see also Altenberg et al., 2018). Beyond the previously discussed measures of relative complexity, we argue that relative complexity can also be operationalized in terms of how effortful it is to process a well-formed structure (inferred by evaluating its associated reading, listening, or looking times, as well as whether it elicits any comprehension errors) and in terms of how effortful it is to produce a well-formed structure (inferred by the presence of disfluencies and speech errors). To a great extent, measures of processing and production difficulty have already been utilized to test psycholinguistic theories of sentence production and comprehension. This has often taken the form of comparisons of grammatical structures that are considered more complex relative to those that are considered less complex (e.g., comprehension of object vs. subject relative clauses: King & Just, 1991; Traxler et al., 2002; production of passive vs. active sentences: Mirdamadi & De Jong, 2015). Thus, we suggest that further incorporation of these measures of processing and production effort, whether in the context of a developing or established knowledge system, would nicely augment existing measures of complexity.

In recent years, automated assessments of complexity have been developed (e.g., Bhat & Yoon, 2015; Lu, 2010; McNamara et al., 2014) allowing researchers to examine a wider variety of syntactic features, thereby allowing for a more comprehensive overview of the complexity of a given language sample. However, such measures have largely focused on capturing properties of written production (see Jagaiah et al., 2020; Ortega, 2003; Wolfe-Quintero, 1998). Even when these measures of complexity have been extended to the study of spoken language, they have primarily been applied to the study of monologue, where individuals are provided with a question/prompt to elicit production (e.g., Nippold et al., 2017; Yu & Lowie, 2020). However, as previously noted, typical language use involves multiple speakers interacting with each other. Thus, it is important to assess how well these previously established measures of complexity apply to the study of actual interactions.

In addition to utilizing measures of complexity to capture language learners' proficiency and performance, these same measures can be used and, to some extent, have been used to characterize how other speakers modify the complexity of their production when interacting with L2 interlocutors. Syntax and syntax-related properties of L2 learner–directed speech[1] have been less well studied than acoustic properties (for a review of acoustic properties of L2 learner–directed speech, see Piazza et al., 2022). Early studies focused on syntactic aspects reported that speech directed toward nonnative listeners is generally characterized by shorter utterances, fewer subordinate clauses, and increased repetitions, both within the classroom where the focus is on instructors' productions directed toward students (Chan et al., 2022; Gaeis, 1977; Henzl, 1973, 1979; Milk, 1990) and outside the classroom under more naturalistic, conversational contexts (Arthur et al., 1980; Freed, 1981; Long, 1980, 1981, 1983). Prior work has also shown that reductions in complexity correspond to L2 interlocutors' proficiency (Chan et al., 2022; Freed, 1981; Henzl, 1979).

One thing to note is that much of this previous work examining syntactic modifications in L2 learner–directed speech have utilized absolute measures of complexity, often reporting differences in length and degree of embeddedness or tracking differences in the types of constructions produced. However, speaking to nonnative interlocutors may not always result in linguistic simplification. For example, speakers may need to define unfamiliar words or concepts to nonnative interlocutors, which would result in longer utterances. As such, when discussing potential syntax-related modifications in L2 learner–directed speech, it is important to consider both absolute complexity and relative complexity (i.e., how effortful it is to produce/process a structure), the implications of complexity on both the speaker and the listener, and the communicative context of the interaction.

Challenges of Characterizing Syntax in Actual Interactions

To date, researchers interested in studying syntax in spoken language have primarily relied on the use of decontextualized language tasks (e.g., grammaticality judgments and sentence repetition tasks), which are restrictive and do not involve true communication. These language tasks are used in language assessments that clinicians utilize to diagnose language disorders (e.g., Comprehensive Assessment of Spoken Language; Carrow-Woolfolk, 2017; Clinical Evaluation of Language Fundamentals [CELF]; Wiig et al., 2013). However, it is possible that such elicitations, often in the context of monologue, may not fully capture the behavior and abilities of language users in naturalistic conversation. An alternative approach involves language sample analyses of speech obtained using discourse elicitation tasks (e.g., describe your favorite game/movie) or picture description tasks that are more naturalistic. However, in such instances the issue is that the clinician or the experimenter has limited control on the elicited samples. Furthermore, given the open-ended nature of the elicitation, it is usually difficult to capture how well the elicited productions contribute to communicative success, in part because it can be difficult to define what communicative success is in the context of open-ended conversations. Given this, we suggest that a referential communication task, such as the Diapix task described earlier, provides a middle ground between (near) complete control over speakers' syntactic productions (e.g., sentence repetition and syntactic priming) and little control over speakers' production (e.g., discourse elicitation and picture description). The Diapix task may be well worth exploring to augment current measures of syntax, since it allows for more control of the semantic content of speakers' productions, making it easier to compare across speakers/groups, and provides researchers with more objective ways of defining and assessing communicative success (e.g., task completion time and number of differences identified).

[1]The speech accommodation literature typically uses the term foreigner-directed speech to describe the register of speech produced when communicating with nonnative speakers of a language. In this research note, we utilize the term L2 learner–directed speech in lieu of foreigner-directed speech to avoid the connotation that something is "alien" or "strange."

More broadly, doing so using a language-as-action approach will help address the problem of studying syntax in more naturalistic production.

At the syntactic level, operating within the language-as-action framework necessitates ceding some control over speakers' speech, allowing for spontaneous production. This presents a challenge because spontaneous speech is often messy. Speakers usually produce filler words, false starts, and other disfluencies. Additionally, speakers sometimes repair/revise their previously produced utterances, which have implications for listeners' comprehension (Lowder & Ferreira, 2016). These disfluencies can further complicate the transcription and segmentation of spoken language. Additionally, researchers particularly interested in assessing syntactic complexity often rely on metrics that quantify the frequency of certain grammatical structures or calculate some ratio-based value that captures the average length of some grammatical unit (e.g., words, phrases, and clauses) or number of some grammatical unit per utterance produced in a language sample (e.g., Bhat & Yoon, 2015; Hwang et al., 2020; Nippold et al., 2017). In order to do this, spoken language must be accurately transcribed and segmented into some unit of analysis, so that these frequencies and ratios can be calculated. Although various units of analysis for segmenting spoken language have been proposed (e.g., T-unit: an independent clause and any suboridinate or subcalusal structure[s] attached to it; Hunt, 1965, 1970; C-unit: independent grammatical units or answers to questions; Loban, 1976; intonation unit: segment of speech produced under an intonational contour; Chafe, 1988), these methods of segmentation fail to address common characteristics of spoken language such as topicalization, coordinate structures, as well as interruptions and scaffolding (i.e., when speakers finish each other's utterances), which are especially common in interaction (see Foster et al., 2000, for further discussion). Since current widely utilized measures, such as T-units, fail to adequately capture spontaneous production, it is necessary to utilize alternative measures that will enable us to capture syntactic features of spontaneous speech, accounting for typical characteristics of naturalistic production (e.g., disfluencies, interruptions, and scaffolding).

Foster et al.'s (2000) Solution

Foster et al. (2000) advocate for the use of the analysis of speech unit (AS-unit), which they define as "a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clause(s) associated with either" (Foster et al., 2000, p. 365). Conceptually, AS-units are similar to T-units (Hunt, 1965, 1970), but Foster et al. (2000) specifically provide guidance on how to deal with common features of spoken language previously discussed above. AS-units also have an advantage over T-units, because the Diapix task encourages collaboration between speakers and often involves asking questions and receiving answers, which sometimes lead to productions of fragments (e.g., Speaker 1: "What do you have in your picture?" Speaker 2: "A woman carrying a blue purse") that would otherwise be excluded under standard definitions of T-units (Hunt, 1965, 1970). Additionally, Foster et al. (2000) noted that researchers sometimes wish to exclude parts of production data depending on the purpose of their analyses, so they proposed three levels of inclusiveness to accommodate different research questions and various types of production data (see Table 1).

The Current Study

As a first step, we adapt existing measures that are better suited to analyzing monologue and aim to understand whether they capture differences in interaction between interlocutors with differing language backgrounds. By doing so, we pave the way for a richer understanding of these aspects of language use in typical conversation. We characterize syntactic adjustments in dialogue using unscripted production data from the Wildcat Corpus (Van Engen et al., 2010) to assess whether various measures of complexity differentiate different dyad types (native, mixed, and nonnative pairs).

We divided our measures of complexity into three categories: measures of absolute syntactic complexity, measures of syntactic rate, and measures of disfluencies.

Table 1. Descriptions of each level of analysis adapted from Foster et al. (2000).

Level of analysis	Application	Exclusion criteria
Level 1	Full analysis of data	Includes all AS-units
Level 2	Analysis of highly interactional data	Excludes "minor one-word utterances" such as backchannels (e.g., "yes," "no," "okay," "uhhuh," "right") Excludes echoic responses (i.e., speaker repeats their partner's previous utterance)
Level 3	Analysis of nonfragmentary AS-units	Excludes "minor one-word utterances" Excludes echoic responses Excludes fragments that cannot stand alone

Note. AS-units = analysis of speech units.

Our measures of absolute syntactic complexity (mean length of AS-unit [MLASU] and clausal density [CD]) correspond to common measures of absolute complexity that have been widely used in SLA and SLHS research (see previous discussion of methods of operationalizing syntactic complexity), with higher MLASU and CD considered to be reflective of greater syntactic complexity. We also examined measures of syntactic rate (number of AS-units, clauses, and words per minute), which captures how much speech a speaker/dyad produced, after accounting for overall speaking time. We treat this as a proxy for the amount of information exchanged during the task.

As discussed, complexity may also be captured by the amount of effort required to produce or comprehend a particular structure. Thus, we included measures of disfluencies to serve as an index of relative complexity. Our measures of disfluencies include number of filled pauses, silent pauses, and linguistic mazes (i.e., words or phrases that disrupt the speaker's intended message and flow of speech; Loban, 1976). Disfluencies often arise in response to increased demands on speech planning, with higher cognitive demand linked to increased disfluency production (Bortfeld et al., 2001). Complex sentence structures are known to require greater cognitive resources to plan (MacDonald, 2013) and produce (Santi et al., 2019; Scontras et al., 2015). This suggests that higher disfluency rates could be attributed to the use of complex grammatical structures during speech production, increased utterance planning effort (e.g., speaking in a L2), or a combination of these factors.

However, although we are most interested in disfluencies as a measure of relative complexity, we acknowledge that the production of disfluencies likely reflects various processes beyond utterance planning difficulties and that different types of disfluencies may reflect different processes. For example, although filled pauses have been treated as a marker of utterance planning difficulties (Clark & Fox Tree, 2002; Smith & Clark, 1993), they have also been argued to serve pragmatic and discourse organization functions, including turn-taking functions (Swerts, 1998; Tottie, 2014; see also Kosmala & Crible, 2022). Relatedly, a distinction has been drawn between the functionality of filled pauses and silent pauses, with filled pauses argued to be more listener-oriented and silent pauses argued to be speaker-oriented (Lake et al., 2011; Shriberg et al., 2001, but see Engelhardt et al., 2017), given that filled pauses could be used to facilitate conversation by indicating the speaker is not yet ready to cede the floor or to signal an upcoming delay (Fox Tree, 2001, 2002). We also note that the location of the disfluency (e.g., within-clause vs. between-clause) may reflect different linguistic processes as well (De Jong, 2016; Kahng, 2014; Tavakoli, 2011). Nonetheless, as a first step to examining production in highly interactive contexts, we only focus on the number of disfluencies speakers produce, leaving future research to further explore how differences in disfluency location and type of disfluency relate to complexity in dialogue.

Comparing across our three dyad types, we predicted that speech produced by native pairs would be more complex than nonnative pairs, reflected by higher MLASU and CD. We anticipated that native speakers would exchange information more efficiently, evidenced by greater syntactic rate (i.e., more AS-units, clauses, and words per minute). Because native speakers are speaking in their L1, we also predicted that native interactions would be accompanied by fewer disfluencies compared to nonnative interactions where speakers are speaking in their L2, which is more cognitively demanding. Because mixed pairs contain both a native and nonnative speaker, we expected mixed pairs to fall in between our other two groups.

In addition to exploring how existing measures of complexity apply to the study of dialogue by comparing across our three pair types, we wanted to utilize these measures to explore whether, at the individual speaker level, modifications in complexity in interactive contexts are driven by the speaker's language background (native vs. nonnative), their interlocutor's language background (native vs. nonnative), or a combination of the two. Regarding the effect of the speaker's language background, we predicted a similar pattern of results as the dyad level for the reasons outline above. Regarding the effect of the interlocutor's language background, prior research suggests that speakers reduce the complexity of their speech when communicating with nonnative interlocutors (e.g., Arthur et al., 1980; Henzl, 1973, 1979; Long, 1983). Thus, if speakers are modifying their production for their interlocutor, we would expect that speaking to nonnative listeners would elicit productions that are less syntactically complex, indexed by shorter MLASU and lower CD. Previous analyses of the Wildcat Corpus reported that dyads that included at least one nonnative speaker were less efficient in completing the Diapix than dyads consisting of two native speakers (Van Engen et al., 2010); thus, we expected that speech produced for non native interlocutors would be characterized by slower syntactic rate and increased disfluencies.

Method

Corpus Description

Because we are primarily interested in characterizing syntactic adjustments in more naturalistic language use, we utilized the unscripted production data in the Wildcat Corpus. These consist of productions that were elicited using the Diapix task. During this task, pairs of participants were instructed to work together to identify ten differences between their two pictures. Participants were instructed to work as quickly as possible with a maximum time limit of 20 min. Because our goal was to capture differences in complexity at the dyad and speaker level, we analyze a subset of Diapix data from the Wildcat Corpus, focusing on the available eight native pairs (English–English) and eight mixed pairs (four English–Chinese and four English–Korean). To match the number of pairs from the other two pair types, we selected eight of the 11 nonnative pairs (four Chinese–Chinese and four Korean–Korean). We opted to focus on nonnative pairs who shared the same L1 to eliminate the possibility that underlying syntactic differences between partners' L1s would contribute to potential interlocutor-driven speaker modifications. All pairs completed the Diapix task in English. For the subset of data that we analyzed, the native English speakers (12 men and 12 women) ranged in age from 18 to 33 years old (M = 20.5). The nonnative speakers (12 men and 12 women) ranged in age from 22 to 33 years old (M = 26.33). All nonnative speakers in the Wildcat Corpus were described as having quite high English proficiency as measured by standardized tests (e.g., TOEFL), but not all nonnative speakers reported their TOEFL or other proficiency score.

Segmentation of Production Data

Transcriptions from the Diapix task portion of the Wildcat Corpus are available online (https://groups.linguistics.northwestern.edu/speech_comm_group/wildcat/). These transcriptions were slightly modified to correct for any typos or instances where the first author (C.T.P.) and one other member of the research team agreed that the transcription did not match what was said in the recording. We adopted the guidelines outlined in Foster et al. (2000) to segment the transcriptions into AS-units. More details regarding the segmentation process can be found in Supplemental Material S1. Clauses were coded following guidelines outlined in Nippold et al. (2014). We identified main, adverbial, relative, nominal, infinitive, participial, and gerund clauses. Verbalized disfluencies (filled pauses and linguistic mazes) were placed in parentheses. Silent pauses were extracted using the TextGrids of Diapix conversations from the Wildcat Corpus, in which silent sections between continuous chunks of speech were already annotated. These TextGrids included separate tiers for each speaker in the dyad, allowing us to differentiate between silent pauses that occurred within an individual speaker's speaking turn and those that occurred between two speakers' turns. Using these TextGrids, we calculated the number and duration of these silent sections (i.e., silent pauses) in Praat. The segmenting and coding of the transcriptions were completed by the first author or one trained research assistant and were cross-checked by the other rater.

Given that we are also interested in exploring the extent to which measures of syntactic complexity apply to dialogue, we conducted Levels 1, 2, and 3 analyses to assess which measures are most informative at each level. Following the guidelines described in Foster et al. (2000), we segmented the transcriptions to accommodate all three levels of analysis. See Table 1 for critical details and Supplemental Material S2 for an example transcription segmented and annotated for each level of analysis. For all three levels, transcriptions were analyzed at both the dyad level and the individual speaker level.

Measures of Complexity

Measures of absolute syntactic complexity. MLASU was calculated by dividing the total number of words by total number of AS-units for each speaker. CD was calculated by dividing the total number of clauses (main and subordinate) by total number of AS-units. Verbalized disfluencies (filled pauses and linguistic mazes) were removed before calculating MLASU and CD for both the dyad and speaker level.

Measures of syntactic rate. Following the removal of all filled pauses and linguistic mazes, we computed the total number of AS-units, clauses, and words for each speaker. Then, to account for differences in the amount of time each pair (and each individual speaker) spent talking, we divided the total number of AS-units, clauses, and words by the total amount of time it took for pairs to complete the Diapix task in minutes (for dyad-level analyses) or by the total speaking time in minutes (for speaker-level analyses).

Measures of disfluencies. Disfluencies were divided into two types: pauses and mazes. We distinguish silent pauses and filled pauses ("ah," "eh," "er," "hm," "mm," "uh," "um"; Maclay & Osgood, 1959). We implemented a 180-ms threshold for distinguishing between silent pauses associated with articulation (< 180 ms) and hesitations (≥ 180 ms) in speech production (Sørensen et al., 2021). When calculating number of silent pauses and average silent pause duration, we only included silent pauses greater than or equal to 180 ms in length. We also note that we were careful to distinguish silent pauses from turn-taking behavior. Our calculation for number of silent pauses and average silent pause duration only includes pauses that occur *within* a speaker's speaking turn, not pauses that occur *between* turns.

We also examined linguistic mazes, which include words or phrases that disrupt the speaker's intended message and flow of speech (Loban, 1976). Researchers sometimes vary in how they define and categorize mazes (Bangert & Finestack, 2020), but we use the term maze to refer to abandoned words/utterances (e.g., "**I wonder-**" or "**banan-**"), false starts/reformulations (the speaker revises their utterance; e.g., "**it says** the sign says Pete's pet shop"), and repetitions (the speaker reproduces the word/phrase verbatim; e.g., "**it's a** it's a sheep"). Note that when the speaker's conversational partner interrupted the speaker, leading to the production of abandoned words/utterances, false starts/reformulations, and repetitions, these instances were not marked as mazes.

Results

Dyad-Level Analyses

For our dyad-level analyses, we calculated a proportion value for number of AS-units, clauses, words, filled pauses, silent pauses, and mazes by dividing these measures by the total duration of the Diapix task to account for differences in the amount of time each pair needed to complete the task. To verify that there were inherent differences between the three types of speaker-interlocutor pairings, we conducted one-way between-subjects analyses of variance (ANOVAs) to examine the effect of Pair Type (native, mixed, and nonnative) on each complexity metric.

Analyses of MLASU, CD, and number of AS-units, clauses, and words per minute were conducted for Levels 1, 2, and 3 transcriptions. Level 1 analyses included all language production data without any exclusions, meaning that all AS-units were included in Level 1 analyses. Level 2 analyses excluded AS-units consisting of echoic responses and backchannels. Level 3 analyses included exclusions applied at Level 2 as well as further exclusions of any AS-units that were sentence fragments (i.e., those that did not contain a main [independent] clause). Because the analyses for Levels 1, 2, and 3 revealed a nearly identical pattern of results, we only present Level 2 results below. (See Supplemental Material S3 for full analyses [Levels 1–3] and further discussion of minor differences across each level.) We opted to focus on Level 2 analyses in accordance Foster et al.'s (2000) suggestion that Level 2 is best applied to highly interactional data, which encompasses the type of production data elicited by the Diapix task. Note that, for our analyses, measures of disfluencies (number of filled pauses, silent pauses, and mazes per minute as well as average duration of silent pauses) do not belong to any particular level of analysis, since the exclusion of AS-units did not affect our overall count of disfluencies. Therefore, we report the results for our measures of disfluencies alongside our Level 2 results.

Measures of Absolute Syntactic Complexity

The top row of Figure 1 shows that native pairs produced the longest AS-units and the greatest number of clauses per AS-units, nonnative pairs produced the shortest AS-units and fewest clauses per AS-units, and mixed pairs fell in between the native and nonnative groups. The data were submitted to a one-way between-subject ANOVA to evaluate whether observed differences between pair types were significant. We observed a main effect of Pair Type on MLASU, $F(2, 21) = 8.28$, $p = .004$, $\eta_p^2 = .44$, and CD, $F(2, 21) = 11.75$, $p < .001$, $\eta_p^2 = .53$. Post hoc comparisons using Tukey's honestly significant difference (HSD) test revealed that group differences in MLASU and CDs were driven by differences between native and nonnative pairs and between mixed and nonnative pairs, with native pairs and mixed pairs producing longer and more clausally dense AS-units than nonnative pairs. Overall, the presence of at least one native speaker in a dyad led to more syntactically complex speech at the dyad level. This pattern of results was expected given that native speakers were expected to be more proficient in English than the nonnative speakers, reflected by higher MLASU and CD. Since mixed pairs consist of both a native and nonnative speaker, it is unsurprising that MLASU and CD of dialogue produced by mixed pairs fell in between the native and nonnative group.

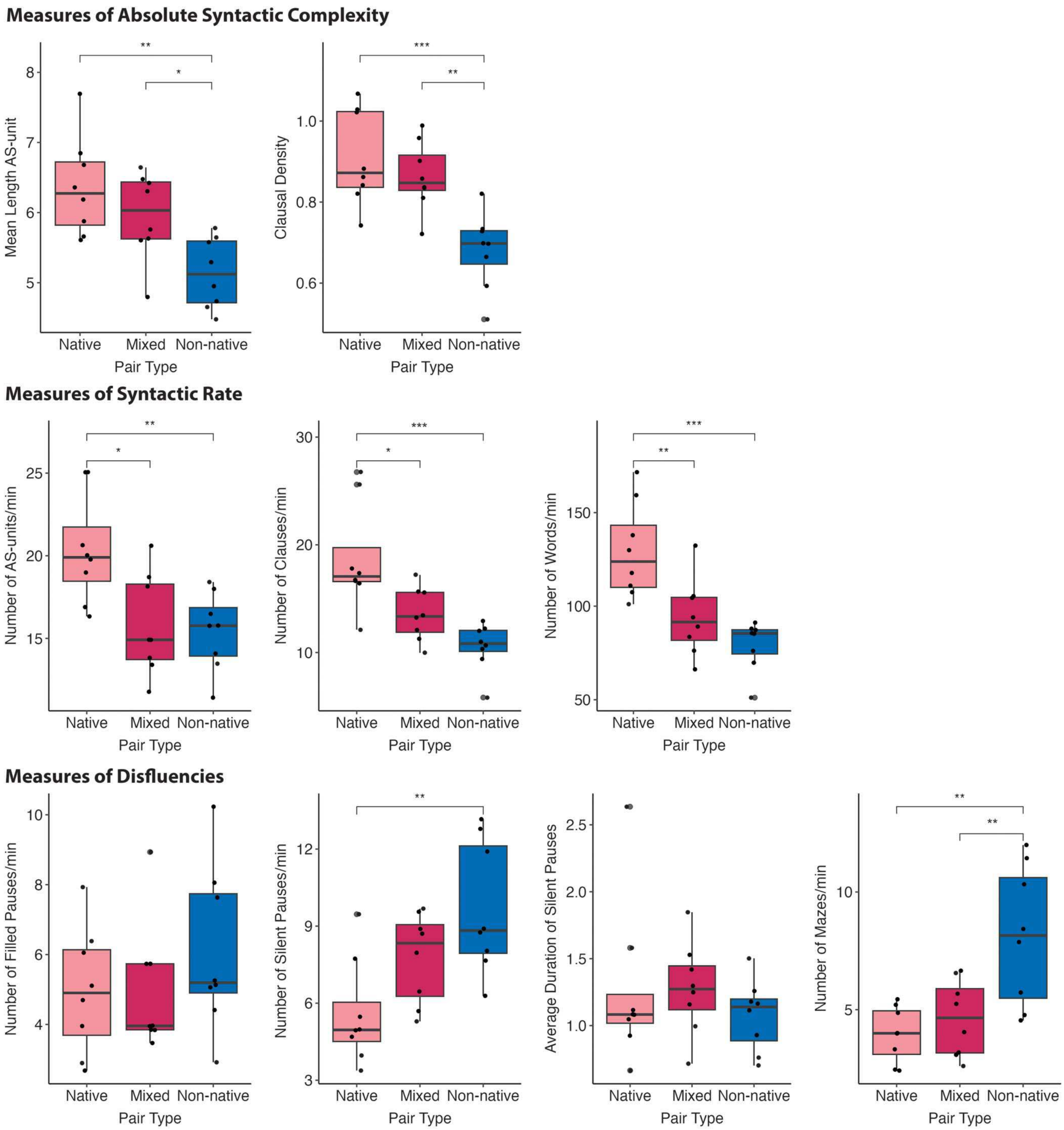
Measures of Syntactic Rate

When examining measures of syntactic rate across our three groups, the middle row of Figure 1 indicates that native pairs produced more speech per minute overall with mixed and nonnative pairs producing similar amounts of speech. Statistical analyses revealed a significant effect of Pair Type on number of AS-units per minute, $F(2, 21) = 7.02$, $p = .004$, $\eta_p^2 = .40$; clauses per minute, $F(2, 21) = 11.48$, $p < .001$, $\eta_p^2 = .52$; and words per minute, $F(2, 21) = 12.92$, $p < .001$, $\eta_p^2 = .55$. Post hoc Tukey's HSD tests revealed that the locus of this difference was between the native and mixed pairs and between the native and nonnative pairs, with native pairs producing more AS-units, clauses, and words per minute than the other two groups, which was expected given that native speakers are expected to be more fluent speakers since they were completing the task in their L1. These findings indicate that, after accounting for differences in total duration of the Diapix task, the presence of a nonnative speaker in a dyad led to a reduction in the amount of speech produced and, consequently, the amount of information exchanged.

Measures of Disfluencies

Visual inspection of the bottom row of Figure 1 shows that nonnative pairs produced the greatest number of filled pauses, silent pauses, and mazes per minute. Non native pairs also produce the shortest average duration of silent pauses. Statistical analyses revealed a main effect of

Figure 1. Boxplots depicting differences between each Pair Type for each complexity measure for Level 2 analyses. Each dot represents one pair of participants. Medians are represented by the horizontal lines. For complexity metrics that exhibited a main effect of Pair Type ($p < .05$), significance bars depict results from post hoc Tukey's honestly significant difference (HSD) tests. AS-unit = analysis of speech unit. *$p < .05$. **$p \leq .01$. ***$p \leq .001$.



Pair Type for number of silent pauses, $F(2, 21) = 7.36$, $p = .004$, $\eta_p^2 = .41$, and mazes per minute, $F(2, 21) = 9.54$, $p = .001$, $\eta_p^2 = .48$. Post hoc analyses using Tukey's HSD test revealed that number of silent pauses and mazes per minute were significantly higher for nonnative pairs compared to native pairs. Nonnative pairs also produced significantly more mazes per minute compared to mixed pairs. The observed increase in silent pauses and mazes for nonnative pairs was expected and presumably reflects utterance planning difficulties, which would be expected to be greater for nonnative speakers, given that they were completing the task in their L2. However, the number of filled pauses per minute did not differ across the three groups ($p < .05$), which was unexpected. Overall, at the dyad level, native pairs produced more syntactically complex speech—evidenced by higher MLASU and CD—a greater quantity of speech per minute, and fewer silent pauses and mazes compared to nonnative pairs, with mixed pairs falling somewhere in between.

Speaker-Level Analyses

For our speaker-level analyses, we calculated a proportion value for number of AS-units, clauses, words, filled pauses, silent pauses, and mazes by dividing these measures by each individual participant's speaking duration to account for differences in speakers' conversational contribution. We conducted two-way between-subject ANOVAs to examine the effect of Speaker Type (native and nonnative), Interlocutor Type (native, nonnative), and the interaction between Speaker Type and Interlocutor Type on each complexity metric.

At the speaker level, analyses of MLASU, CD, and number of AS-units, clauses, and words per minute were conducted for Levels 1, 2, and 3 transcriptions. We also conducted analyses for our measures of disfluencies (number of filled pauses, silent pauses, and mazes per minute and average duration of silent pauses), but we do not attribute measures of disfluencies to any of our levels (1–3) of analyses since the exclusion of AS-units did not affect our overall count of disfluencies. Once again, analyses for Levels 1, 2, and 3 revealed a nearly identical pattern of results, so we only present Level 2 results below. (See Supplemental Material S3 for full analyses [Levels 1–3] and further discussion of minor differences across each level.)

Measures of Absolute Syntactic Complexity

The top row of Figure 2 shows that native speakers produced shorter AS-units and fewer clauses per AS-units compared to nonnative speakers. The data were submitted to a two-way between-subject ANOVA to confirm this pattern of results. A significant main effect of Speaker Type was observed for MLASU, $F(1, 44) = 7.78$, $p = .008$, $\eta_p^2 = .15$. As expected, native English speakers produced significantly longer AS-units compared to nonnative English speakers. Additionally, although native speakers did produce more clauses per AS-units than nonnative speakers, this effect was marginal, $F(1, 44) = 3.88$, $p = .06$, $\eta_p^2 = .08$. Interlocutor Type did not have an effect on MLASU or CD. Although we expected that speaking to nonnative speakers would lead to reductions in syntactic complexity, we failed to detect any differences on these metrics as a function of Interlocutor Type. Although visual inspection of Figure 2 suggests a potential interaction of Speaker Type and Interlocutor Type on CD, there was no significant interaction between Speaker Type and Interlocutor Type on CD (nor MLASU). Thus, the complexity of a speaker's production was dependent on their own language background (native vs. nonnative), and speakers did not seem to modify the length or CD of their AS-units based on their interlocutor's language background.

Measures of Syntactic Rate

At the individual speaker level, the middle row of Figure 2 suggests the amount of speech produced per minute was determined by the speaker's language background, with native speakers producing more AS-units, clauses, and words per minute relative to nonnative speakers. This was confirmed by an observed main effect of Speaker Type for the number of AS-units, $F(1, 44) = 13.63$, $p = .001$, $\eta_p^2 = .24$; clauses, $F(1, 44) = 16.86$, $p = < .001$, $\eta_p^2 = .28$; and words, $F(1, 44) = 29.19$, $p < .001$, $\eta_p^2 = .37$, per minute. Additionally, the main effect of Interlocutor Type and the interaction between Speaker Type and Interlocutor type were not significant for any of our measures of syntactic rate. Overall, the speaker's language background determined their syntactic rate, and speakers did not adjust how quickly they spoke based on their interlocutor's language background.

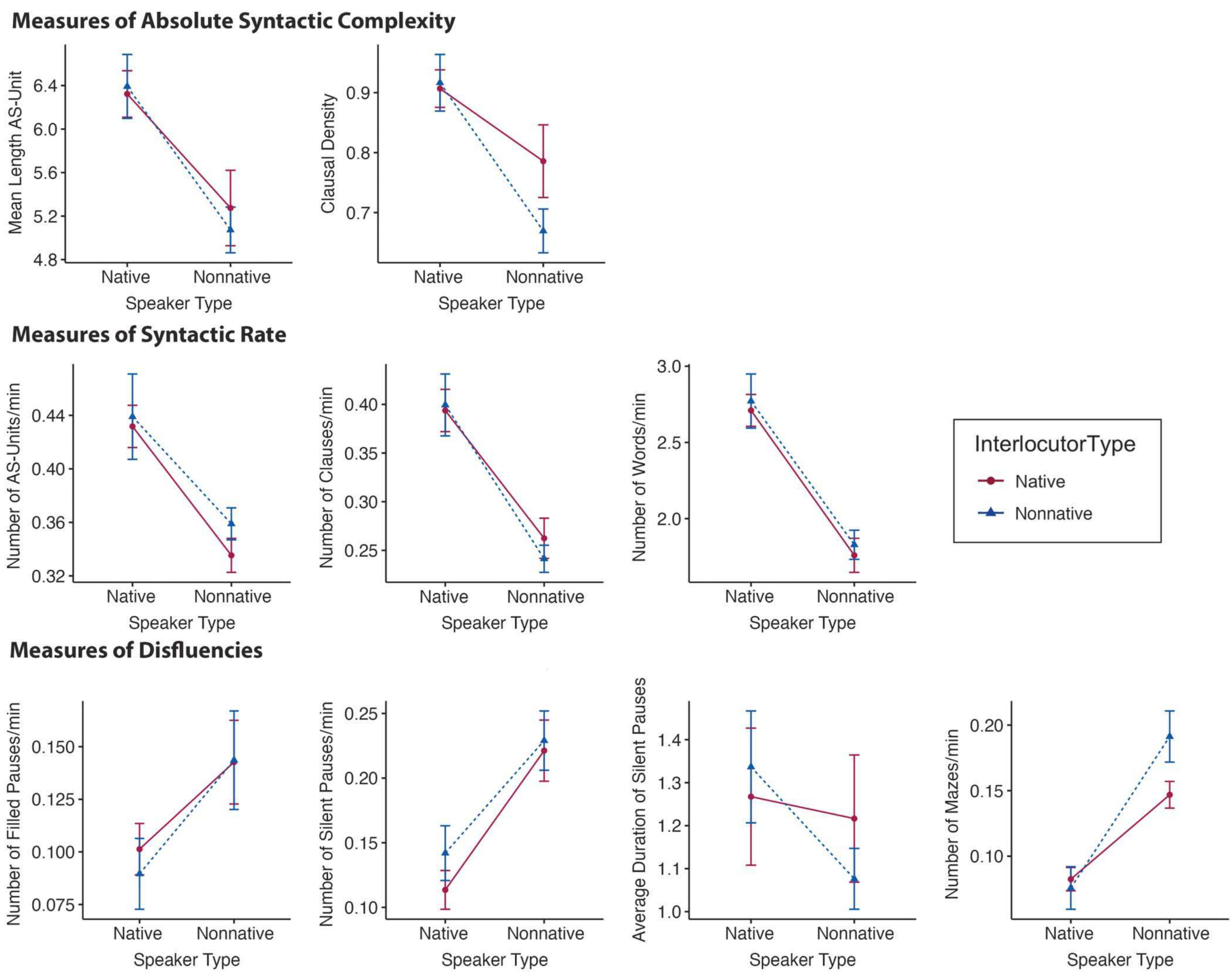
Measures of Disfluencies

The bottom row of Figure 2 illustrates that native speakers produced fewer filled pauses, fewer silent pauses, and fewer mazes per minute compared to nonnative speakers. Inferential analyses revealed that differences in the number of filled pauses produced did not significantly differ between native and nonnative speakers, $F(1, 44) = 1.95$, $p = .16$, $\eta_p^2 = .04$. Although we did observe a significant main effect of Speaker Type on number of mazes, $F(1, 44) = 7.43$, $p = .009$, $\eta_p^2 = .14$, and silent pauses per minute, $F(1, 44) = 11.49$, $p = .001$, $\eta_p^2 = .21$, the effect of Speaker Type was not significant for average silent pause duration, $F(1, 44) = 0.06$, $p = .80$, $\eta_p^2 = .001$. The main effect of Interlocutor Type and the interaction between Speaker Type and Interlocutor Type were not significant for any of our measures of disfluencies. Taken together, these results suggest that the number of silent pauses and mazes produced per minute was again dependent on the speaker's language background and not on their interlocutor's language background.

Discussion

The current study examined speaker modifications at the syntactic level. Unlike prior work exploring

Figure 2. Interaction plots depicting Level 2 analyses for the effect of Speaker Type and Interlocutor Type on each complexity metric computed. Error bars represent standard errors. AS-unit = analysis of speech unit.



modifications in syntactic complexity, we examined how speakers may modify syntactic aspects of their language production in interaction. Overall, speakers' productions reflected their status as native or nonnative speakers of English. Native speakers generally produced longer and more clausally dense AS-units, a greater quantity of speech, and fewer silent pauses and linguistic mazes than nonnative speakers. This was true at both the dyad level and the individual speaker level. These patterns of results are largely expected and serve as confirmation that our various measures can be successfully applied to the study of dialogue.

Interestingly, these group differences did not manifest when examining the production of filled pauses, which was surprising given that we expected nonnative speakers to produce more filled pauses in line with previous literature (Götz, 2013). Prior work has also reported that nonnative speakers tend to produce more filled pauses within clauses than between clauses (De Jong, 2016; Tavakoli, 2011). Because we did not differentiate between the location of filled pauses, that might be one potential reason we did not observe a difference between natives and non natives in the current study. A similar argument in favor of differentiating location of occurrence could be made for silent pauses (see Kahng, 2014). Thus, in the future, more careful consideration should be given to pause location when studying complexity.

Previous studies have also reported that native speakers produce more filled pauses while nonnative speakers tend to produce more silent pauses (Cenoz, 2000;

Iwashita et al., 2008). In our study, nonnative speakers produced more of both types of pauses relative to native speakers, although the effect for filled pauses was not significant. One possibility is that nonnative speakers produced more filled pauses than native speakers because the original intent behind Van Engen et al.'s (2010) Diapix images was to elicit the production of speech sounds that were meant to be difficult for nonnative speakers. As such speakers may have produce more filled pauses before difficult to produce lexical item. Alternatively, nonnative speakers may have heavily relied on filled pauses to maintain the floor during the Diapix, leading to greater production of filled pauses. Furthermore, more nuanced investigation of filled pause location is needed to tease these two possibilities apart.

Overall, when examining the speaker's linguistic status (native vs. nonnative), we observed similarities as well as meaningful differences with previous studies that did not use communicative tasks to elicit production Since we utilized a communicative task in, we can also investigate how the interlocutor's linguistic status (native vs. nonnative) may have impacted the complexity of the speaker's production. The idea that language users take into account the identity and knowledge of their conversation partner is referred to as audience design. Audience design proposes that speakers modify their production to accommodate the needs of their interlocutor (Clark & Murphy, 1982). Thus, from an audience design perspective, speakers should produce less syntactically complex speech to facilitate their interlocutor's comprehension. However, our results unexpectedly showed that the language background of the interlocutor did not impact the syntactic complexity of speakers' AS-units nor the amount of speech or number of disfluencies produced. In other words, speakers did not modify syntactic aspects of production to accommodate their conversation partner. Our findings contradict prior research reporting that speech directed toward nonnative interlocutors was less syntactically complex—characterized by shorter utterances, decreased subordinate clauses and filled pauses, and increased repetitions and silent pauses (e.g., Arthur et al., 1980; Henzl, 1973, 1979; Long, 1983).

One possible interpretation of these results is that listener-oriented adaptations occur primarily at the acoustic–phonetic level and less so at the syntactic level.[2] Thus, speakers may only alter the syntactic aspects of their production for their own benefit as opposed to for their interlocutor's benefit. This interpretation would be in accordance with extant literature examining audience design at the syntactic level. Prior work in this domain has primarily focused on exploring whether speakers avoid producing temporary syntactic ambiguities that are known to lead to processing difficulties (i.e., garden path sentences). Overall, these studies suggest that speakers do not engage in ambiguity avoidance at the syntactic level (Arnold et al., 2004; Ferreira & Dell, 2000; Ferreira & Schotter, 2013; Jaeger, 2010; but see Temperley, 2003), which has been taken as evidence that speakers' syntactic choices are speaker driven, not listener driven (see Arnold, 2011). However, it is important to note that these studies focus on the production of very specific (and arguably very complex) syntactic structures, making it difficult to generalize these findings to more naturalistic language production such as the current study. In addition, it may also be that acoustic–phonetic adjustments are easier to detect given the graded nature of their measurement, rather than, for example, at the level of syntax.

This is not to say that speakers in the current study did not modify their production in response to communicative demands. One striking difference observed was the difference in the duration of the interactions between the native pairs and the pairs containing nonnative speakers. Interactions involving nonnative speakers were longer, which may be attributed to speakers needing to repeat themselves, paraphrase, and/or define unfamiliar vocabulary words when speaking to nonnative interlocutors. This behavior is a form of elaboration that may be linguistically more complex on the surface but cognitively simpler (i.e., easier) for the listener.

It is possible that we did not observe evidence of interlocutor-driven syntactic modifications due to the task dynamics of the interactive task chosen for this study. The images utilized in the Diapix task by Van Engen et al. (2010) were originally created to probe acoustic–phonetic speaker adjustments. As such, the differences between the two images consist of differences where an item is present in one image but absent from the other or where an item is slightly different between the two images (e.g., different colors). Given that speakers were trying to identify as many differences as possible, they could have adopted a strategy to use shorter and less clausally dense utterances to streamline their conversations, which would be a particularly effective strategy given that the differences could be identified by simply naming various objects, colors, and signs. Thus, a consideration for future research is the types of differences depicted in the Diapix images. To further assess speaker modifications at the syntactic level, it would be beneficial to utilize images where differences between images center on differing actions being performed by characters in the scene (e.g., a dog chasing a cat vs. a cat chasing a dog) as opposed to simply manipulating objects in the scene to encourage speakers to produce more complex structures. Additionally, future research could also explore how the use of other interactive tasks (e.g., map

[2]We extend thanks to an anonymous reviewer for this suggestion.

task) may alter the observed patterns in speakers' modifications. This would be expected under interpersonal synergy accounts of dialogue, which argue that task dynamics are likely to affect both conversation partners' behavior at different levels of analyses (Fusaroli et al., 2014; Olmstead et al., 2021).

One consideration when utilizing goal-oriented tasks such as the Diapix to examine syntax is that the nature of the task may limit the variety of syntactic structures that speakers produce, potentially contributing to the observed lack of syntactic accommodation. However, significant evidence also exists for syntactic alignment in tasks that are more constrained than the Diapix (e.g., Branigan et al., 2000; Ivanova et al., 2020). One reason why we potentially did not observe evidence of syntactic accommodation in the current study is that complexity measures do not provide any information about the types of structures speakers produced. Thus, a more qualitative assessment of the production data may reveal differences in syntactic complexity not captured here. Future analyses of the data should consider potential differences in the overall patterns in the types of grammatical constructions speakers produce and how these patterns might shift depending on interlocutor identity.

Finally, a notable limitation of the current work is that our complexity measures do not capture the dynamics of speakers' productions during the task. It would be helpful to examine measures of complexity in tandem with adjustments in turn-taking dynamics, which have previously been linked to increased communicative effort (Sørensen et al., 2021), and the communicative efficiency of the task. We are presently pursuing further research aimed at exploring the link between complexity and efficiency. Future investigations could also quantify complexity incrementally. For example, information-theoretic measures such as surprisal (the predictability of a word given its preceding context; Hale, 2001; Levy, 2008) and entropy reduction (the amount of information gained after encountering a word; Hale, 2003, 2006) are incremental complexity metrics that can be calculated for each word in a sentence. However, surprisal and entropy reduction can also be measured at the sentence level by averaging the surprisal/entropy values of each word in the sentence. Incorporating these types of measures would allow researchers to examine how the complexity of speakers' production may evolve throughout the conversation.

Nevertheless, in order to better understand naturalistic language use, it is imperative that language scientists undertake the study of interactive tasks. The present investigation represents a critical step in that direction, establishing the groundwork for future research that delves into the intricacies of speaker modifications of complexity in interactive settings. Additionally, models of production and comprehension are largely based on native-native interactions, so it is likely these models fail to incorporate a broad range of speaker adjustments in conversation. Our study provides modest steps toward broadening the study of interaction in different languages across speakers with different linguistic profiles.

Data Availability Statement

The data sets generated and analyzed during the current study as well as analysis scripts are publicly available at https://osf.io/t7nzc/.

Acknowledgments

This research was supported by National Science Foundation Grant B2-2126888, awarded to Navin Viswanathan. The authors thank Olivia Billetdeux, Wayne Kim, Lauren Schwartz, Vacha Shah, Yuka Tatsumi, and Chloe Webb for their assistance transcribing and annotating the Wildcat production data.

References

Altenberg, E. P., Roberts, J. A., & Scarborough, H. S. (2018). Young children's structure production: A revision of the Index of Productive Syntax. *Language, Speech, and Hearing Services in Schools, 49*(4), 995–1008. https://doi.org/10.1044/2018_lshss-17-0092

Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., & Weinert, R. (1991). The Hcrc map task corpus. *Language and Speech, 34*(4), 351–366. https://doi.org/10.1177/002383099103400404

Arnold, J. E. (2011). Ordering choices in production: For the speaker or for the listener? In E. M. Bender & J. E. Arnold (Eds.), *Language from a cognitive perspective: Grammar, usage, and processing: Studies in honor of Thomas Wasow* (pp. 199–222). CSLI Publications.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language, 51*(1), 55–70. https://doi.org/10.1016/j.jml.2004.03.006

Arthur, B., Weiner, R., Culver, M., Lee, Y. J., & Thomas, D. (1980). The register of impersonal discourse to foreigners: Verbal adjustments to foreign accent. In D. Larsen-Freeman (Ed.), *Discourse analysis in second language research* (pp. 111–124). Newbury House Publishers.

Baker, R., & Hazan, V. (2011). DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods, 43*(3), 761–770. https://doi.org/10.3758/s13428-011-0075-y

Bangert, K. J., & Finestack, L. H. (2020). Linguistic maze production by children and adolescents with attention-deficit/hyperactivity disorder. *Journal of Speech, Language, and*

Hearing Research, 63(1), 274–285. https://doi.org/10.1044/2019_JSLHR-19-00187

Bhat, S., & Yoon, S.-Y. (2015). Automatic assessment of syntactic complexity for spontaneous speech scoring. *Speech Communication, 67,* 42–57. https://doi.org/10.1016/j.specom.2014.09.005

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech, 44*(2), 123–147. https://doi.org/10.1177/00238309010440020101

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13–B25. https://doi.org/10.1016/S0010-0277(99)00081-5

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins. https://doi.org/10.1075/lllt.32.02bul

Carrow-Woolfolk, E. (2017). *CASL-2: Comprehensive Assessment of Spoken Language*. Western Psychological Services.

Cenoz, J. (2000). Pauses and hesitation phenomena in second language production. *ITL - International Journal of Applied Linguistics, 127*(1), 53–69. https://doi.org/10.1075/itl.127-128.03cen

Chafe, W. L. (1988). Linking intonation units in spoken English. In J. Halman & S. A. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 1–27). John Benjamins. https://doi.org/10.1075/tsl.18.03cha

Chan, K. C. J., Monaghan, P., & Michel, M. (2022). Adapting to children's individual language proficiency: An observational study of preschool teacher talk addressing monolinguals and children learning English as an additional language. *Journal of Child Language, 50*(2), 365–390. https://doi.org/10.1017/S0305000921000854

Clark, H. H. (1992). *Arenas of language use*. University of Chicago Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition, 84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology, 9,* 287–299. https://doi.org/10.1016/S0166-4115(09)60059-5

De Jong, N. H. (2016). Predicting pauses in L1 and L2 speech: The effects of utterance boundaries and word frequency. *International Review of Applied Linguistics in Language Teaching, 54*(2), 113–132. https://doi.org/10.1515/iral-2016-9993

Engelhardt, P. E., Alfridijanta, O., McMullon, M. E. G., & Corley, M. (2017). Speaker-versus listener-oriented disfluency: A re-examination of arguments and assumptions from autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*(9), 2885–2898. https://doi.org/10.1007/s10803-017-3215-0

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40*(4), 296–340. https://doi.org/10.1006/cogp.1999.0730

Ferreira, V. S., & Schotter, E. R. (2013). Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors? *Quarterly Journal of Experimental Psychology, 66*(8), 1548–1571. https://doi.org/10.1080/17470218.2012.753924

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354–375. https://doi.org/10.1093/applin/21.3.354

Fox Tree, J. E. (2001). Listeners' uses of *um* and *uh* in speech comprehension. *Memory & Cognition, 29*(2), 320–326. https://doi.org/10.3758/bf03194926

Fox Tree, J. E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes, 34*(1), 37–55. https://doi.org/10.1207/S15326950DP3401_2

Freed, B. F. (1981). Foreigner talk, baby talk, native talk. *International Journal of the Sociology of Language, 1981*(28), 19–39. https://doi.org/10.1515/ijsl.1981.28.19

Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology, 32,* 147–157. https://doi.org/10.1016/j.newideapsych.2013.03.005

Gaeis, S. (1977). The nature of linguistic input in formal language learning: Linguistic and communicative strategies in ESL teachers' classroom language. In H. D. Brown, C. Yorio, & R. H. Crymes (Eds.), *On TESOL '77: Teaching and learning English as a second language: Trends in research and practice* (pp. 204–212). https://eric.ed.gov/?id=ED187115

Götz, S. (2013). *Fluency in native and nonnative English speech*. John Benjamins. https://doi.org/10.1075/scl.53

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics, 2,* 159–166. https://www.aclweb.org/anthology/N01-1021

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research, 32*(2), 101–123. https://doi.org/10.1023/A:1022492123056

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30*(4), 643–672. https://doi.org/10.1207/s15516709cog0000_64

Hazan, V., & Baker, R. (2011). Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustical Society of America, 130*(4), 2139–2152. https://doi.org/10.1121/1.3623753

Henzl, V. M. (1973). Linguistic register of foreign language instruction. *Language Learning, 23*(2), 207–222. https://doi.org/10.1111/j.1467-1770.1973.tb00656.x

Henzl, V. M. (1979). Foreign talk in the classroom. *International Review of Applied Linguistics, 17*(2), 159–167. https://doi.org/10.1515/iral.1979.17.1-4.159

Housen, A., De Clercq, B., Kuiken, F., & Vedder, I. (2019). Multiple approaches to complexity in second language research. *Second Language Research, 35*(1), 3–21. https://doi.org/10.1177/0267658318809765

Hunt, K. W. (1965). *Grammatical structures at three grade levels*. The National Council of Teachers of English Research Report No. 3.

Hunt, K. W. (1970). Syntactic maturity in schoolchildren and adults. *Monographs of the Society for Research in Child Development, 35*(1), 1–67. https://doi.org/10.2307/1165818

Hwang, H., Jung, H., & Kim, H. (2020). Effects of written versus spoken production modalities on syntactic complexity measures in beginning-level child EFL learners. *The Modern Language Journal, 104*(1), 267–283. https://doi.org/10.1111/modl.12626

Ivanova, I., Horton, W. S., Swets, B., Kleinman, D., & Ferreira, V. S. (2020). Structural alignment in dialogue and monologue (and what attention may have to do with it). *Journal of Memory and Language, 110*. https://doi.org/10.1016/j.jml.2019.104052

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24–49. https://doi.org/10.1093/applin/amm017

Iyer, N., Thompson, E., Stillwagon, K., Ennis, Z., Willis, A., & Simpson, B. (2016). Adaptive speech modifications and its effect on communication effectiveness in complex acoustic environments. *The Journal of the Acoustical Society of America, 140*(Suppl. 4.), Article 3437. https://doi.org/10.1121/1.4971074

Jaeger, F. T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*(1), 23–62. https://doi.org/10.1016/j.cogpsych.2010.02.002

Jagaiah, T., Olinghouse, N. G., & Kearns, D. M. (2020). Syntactic complexity measures: Variation by genre, grade-level, students' writing abilities, and writing quality. *Reading and Writing, 33*(10), 2577–2638. https://doi.org/10.1007/s11145-020-10057-x

Kahng, J. (2014). Exploring utterance and cognitive fluency of L1 and L2 English speakers: Temporal measures and stimulated recall. *Language Learning, 64*(4), 809–854. https://doi.org/10.1111/lang.12084

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*(5), 580–602. https://doi.org/10.1016/0749-596X(91)90027-H

Kosmala, L., & Crible, L. (2022). The dual status of filled pauses: Evidence from genre, proficiency and co-occurrence. *Language and Speech, 65*(1), 216–239. https://doi.org/10.1177/00238309211010862

Lake, J. K., Humphreys, K. R., & Cardy, S. (2011). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin & Review, 18*(1), 135–140. https://doi.org/10.3758/s13423-010-0037-x

Lee, D.-Y., & Baese-Berk, M. M. (2020). The maintenance of clear speech in naturalistic conversations. *The Journal of the Acoustical Society of America, 147*(5), 3702–3711. https://doi.org/10.1121/10.0001315

Lee, L. L., & Canter, S. M. (1971). Developmental sentence scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders, 36*(3), 315–340. https://doi.org/10.1044/jshd.3603.315

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177. https://doi.org/10.1016/j.cognition.2007.05.006

Loban, W. (1976). *Language development: Kindergarten through grade twelve.* The National Council of Teachers of English Research Report No. 18. https://eric.ed.gov/?id=ED128818

Long, M. H. (1980). *Input, interaction, and second language acquisition.* [Doctoral dissertation, University of California]. https://www.proquest.com/docview/303009869/abstract/4DE60D57886949A7PQ/1

Long, M. H. (1981). Questions in foreigner talk discourse. *Language Learning, 31*(1), 135–157. https://doi.org/10.1111/j.1467-1770.1981.tb01376.x

Long, M. H. (1983). Native speaker/non-native speaker conversation and the negotiation of comprehensible input. *Applied Linguistics, 4*(2), 126–141. https://doi.org/10.1093/applin/4.2.126

Lowder, M. W., & Ferreira, F. (2016). Prediction in the processing of repair disfluencies. *Language, Cognition and Neuroscience, 31*(1), 73–79. https://doi.org/10.1080/23273798.2015.1036089

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4.* https://doi.org/10.3389/fpsyg.2013.00226

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *WORD, 15*(1), 19–44. https://doi.org/10.1080/00437956.1959.11659682

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press. https://doi.org/10.1017/CBO9780511894664

Miestamo, M. (2008). Grammatical complexity in cross-linguistic perspective. In M. Miestamo, K. Sinnemäki, & F. Karlsson (Eds.), *Language complexity: Typology, contact, change* (pp. 23–41). John Benjamins. https://doi.org/10.1075/slcs.94.04mie

Miestamo, M. (2009). Implicational hierarchies and grammatical complexity. In G. Sampson, D. Gil, & P. Trudgill (Eds.), *Language complexity as an evolving variable* (pp. 80–97). Oxford University Press.

Milk, R. D. (1990). Can foreigners do "foreigner talk"?: A study of the linguistic input provided by non-native teachers of EFL. *Texas Papers in Foreign Language Education, 1*(4), 274–288.

Mirdamadi, F. S., & De Jong, N. H. (2015). The effect of syntactic complexity on fluency: Comparing actives and passives in L1 and L2 speech. *Second Language Research, 31*(1), 105–116. https://doi.org/10.1177/0267658314554498

Nippold, M. A., Cramond, P. M., & Hayward-Mayhew, C. (2014). Spoken language production in adults: Examining age-related differences in syntactic complexity. *Clinical Linguistics & Phonetics, 28*(3), 195–207. https://doi.org/10.3109/02699206.2013.841292

Nippold, M. A., Frantz-Kaspar, M. W., & Vigeland, L. M. (2017). Spoken language production in young adults: Examining syntactic complexity. *Journal of Speech, Language, and Hearing Research, 60*(5), 1339–1347. https://doi.org/10.1044/2016_JSLHR-L-16-0124

Olmstead, A. J., Viswanathan, N., Cowan, T., & Yang, K. (2021). Phonetic adaptation in interlocutors with mismatched language backgrounds: A case for a phonetic synergy account. *Journal of Phonetics, 87.* https://doi.org/10.1016/j.wocn.2021.101054

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America, 119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Piazza, G., Martin, C. D., & Kalashnikova, M. (2022). The acoustic features and didactic function of foreigner-directed speech: A scoping review. *Journal of Speech, Language, and Hearing Research, 65*(8), 2896–2918. https://doi.org/10.1044/2022_JSLHR-21-00609

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(02), 169–190. https://doi.org/10.1017/S0140525X04000056

Santi, A., Grillo, N., Molimpakis, E., & Wagner, M. (2019). Processing relative clauses across comprehension and production: Similarities and differences. *Language, Cognition and Neuroscience, 34*(2), 170–189. https://doi.org/10.1080/23273798.2018.1513539

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics, 11*(1), 1–22. https://doi.org/10.1017/S0142716400008262

Scontras, G., Badecker, W., Shank, L., Lim, E., & Fedorenko, E. (2015). Syntactic complexity effects in sentence production. *Cognitive Science, 39*(3), 559–583. https://doi.org/10.1111/cogs.12168

Shriberg, L. D., Rhea, P., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research, 44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*(1), 25–38. https://doi.org/10.1006/jmla.1993.1002

Sørensen, A. J. M., Fereczkowski, M., & MacDonald, E. N. (2021). Effects of noise and second language on conversational dynamics in task dialogue. *Trends in Hearing, 25*. https://doi.org/10.1177/23312165211024482

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics, 30*(4), 485–496. https://doi.org/10.1016/S0378-2166(98)00014-9

Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal, 65*(1), 71–79. https://doi.org/10.1093/elt/ccq020

Temperley, D. (2003). Ambiguity avoidance in English relative clauses. *Language, 79*(3), 464–484. https://doi.org/10.1353/lan.2003.0189

Tottie, G. (2014). On the use of *uh* and *um* in American English. *Functions of Language, 21*(1), 6–29. https://doi.org/10.1075/fol.21.1.02tot

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language, 47*(1), 69–90. https://doi.org/10.1006/jmla.2001.2836

Tuomainen, O., Taschenberger, L., Rosen, S., & Hazan, V. (2021). Speech modifications in interactive speech: Effects of age, sex and noise type. *Philosophical Transactions of the Royal Society B, 377*(1841), Article 20200398. https://doi.org/10.1098/rstb.2020.0398

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech, 53*(4), 510–540. https://doi.org/10.1177/0023830910372495

Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals—Fifth Edition (CELF-5)*. Pearson.

Wolfe-Quintero, K. (1998). *Second language development in writing: Measures of fluency, accuracy, and complexity*. National Foreign Language Resource Center.

Yu, H., & Lowie, W. (2020). Dynamic paths of complexity and accuracy in second language speech: A longitudinal case study of Chinese learners. *Applied Linguistics, 41*(6), 855–877. https://doi.org/10.1093/applin/amz040