Research Methods in Applied Linguistics 3 (2024) 100132

Contents lists available at ScienceDirect

Research Methods in Applied Linguistics

journal homepage: www.elsevier.com/locate/rmal

Coding all clauses in L2 data: A call for consistency

MaryLou Vercellotti [a,*], Sean Hall [a,b]

[a] Ball State University, IN 47306, United States
[b] South China Agricultural University

ARTICLE INFO

Keywords:
Complexity
Finite clauses
Nonfinite clauses
Subordination
Reliability

ABSTRACT

Linguistic complexity is considered a hallmark of language proficiency (Norris & Ortega, 2009), and it is often a focus of second language (L2) research (Pallotti, 2015). Within the construct of complexity, syntactic complexity considers the combination of words and clauses. The term *clause*, however, is underspecified in the field, and L2 researchers only occasionally define how *clause* is operationalized in the coding. After a review of the importance of consistent clause coding, this commentary argues that coordinated verb phrases, nonfinite clauses, and verbal small clauses can reach the complexity of finite clauses by the production of another element, making these constructions commensurate with subordinate finite clauses. Since these constructions are structurally similar to accepted clauses (e.g., imperatives), their exclusion from the clausal complexity calculation threatens the validity and reliability of the coding. Because a narrow definition requires an arbitrary division along a continuum of complex verbal constructions, the field should adopt this broader definition of *clause*. Further, researchers should clarify their methodology choices and acknowledge coding challenges, which are common in L2 data, especially L2 speech data. The increased transparency in research methodology supports the interpretation of research results and facilitates comparisons across L2 studies.

With accuracy and fluency, complexity is often considered a hallmark of second language (L2) proficiency (Norris & Ortega, 2009), and, as a result, complexity is often a focus of L2 research (Pallotti, 2015). To capture syntactic complexity, L2 researchers have employed many different measures (Norris & Ortega, 2009; Pallotti, 2015). Specifically, L2 researchers have often used clause-based measures (e.g., el Majidi, de Graaff & Janssen, 2021; Vercellotti, 2018). Researchers, however, seldom define how *clause* is operationalized in the coding (e.g., Ahmadian, Tavakoli & Dastjerdi, 2015; Leonard & Shea, 2017; Skehan, Bui, Wang & Shum, 2024). For instance, Leonard and Shea (2017: p. 182) described their coding of syntactic complexity as "subordinate clauses were marked," and Ahmadian et al. (2015: p. 48) stated that syntactic complexity was calculated as "the ratio of clauses to [analysis of speech] AS-units." Perhaps the identification of (finite) clauses seems apparent and uncomplicated, so the details were not provided given space limitations.[1] Still, vagueness obscures relevant coding choices, hinders the interpretation of research findings, and impedes meaningful comparisons across research studies.

This paper's first section explains the importance of clearly defining *clause* in language research articles. The second section reviews a continuum of complex verbal constructions, using examples from existing L2 speech data, and justifies which constructions reach clause status. The third and four sections, respectively, review some coding complications in learner data and in speech data, as such data do not always match the forms given in grammar books. The final section summarizes a broader definition of *clause*, which both

* Corresponding author.
E-mail address: mlvercellott@bsu.edu (M. Vercellotti).

[1] In contrast, Vajjala and Meurers (2012) and McManus et al. (2021) thoroughly described clause coding procedures.

https://doi.org/10.1016/j.rmal.2024.100132
Received 16 March 2024; Received in revised form 19 June 2024; Accepted 20 June 2024
Available online 28 June 2024

captures clausal complexity and facilitates the consistent coding of clauses across L2 research studies.

Clause coding choices skew measures of complexity and beyond

Norris and Ortega (2009) stated that generally researchers have used measures of length and measures of subordination to capture syntactic complexity. Some measures, including mean clause length and clauses per sentence, are dependent on what counts as a clause. If *clause* is operationalized more narrowly (e.g., finite verbs with an overt subject), there will be fewer coded clauses, but the average clause length will be higher. Conversely, if a broader definition of clause is chosen (e.g., both finite and nonfinite clauses), the number of clauses will increase while the mean clause length decreases as the number of words will be averaged across a larger number of (smaller) units. The effect on the calculation is obvious, and perhaps seemingly minor, but the size of the difference depends upon the number of constructions that are included or excluded, which is not negligible when looking more closely. For instance, the production of nonfinite clauses has been found to align with proficiency (e.g., Vercellotti, 2018; Vercellotti & Packer, 2016), which means the inclusion or exclusion of nonfinite clauses in the clause-based calculations will be consequential with longitudinal data or data elicited from advanced learners. Furthermore, since many studies (e.g., Bulté & Roothooft, 2020, De Jong & Vercellotti, 2016; el Majidi et al., 2021; Leonard & Shea, 2017; McManus, Mitchell & Tracy-Ventura, 2021; Skehan et al., 2024) have operationalized complexity as a subordination ratio, readers need to understand which subordinate clauses were included in that calculation for results to be interpreted and compared across L2 studies.

Moreover, the clause coding choices can affect other language performance measures. Measures of accuracy can be dependent on the coding of clauses in research studies where accuracy is operationalized as percentage of error-free clauses (e.g., De Jong & Vercellotti, 2016; el Majidi et al., 2021; McManus et al., 2021; Skehan et al., 2024; Vercellotti, 2017). Even measures of fluency can be dependent on the definition of *clause* when the fluency measure considers whether pausing occurs at a clause boundary (e.g., Leonard & Shea, 2017). Hence, the coding of clauses could have implications for researching accuracy and fluency as well as complexity.

Which clauses count?

According to Vajjala and Meurers (2012: 167), "a clause is any structure with a subject and a finite verb." Given this definition, clause coding may seem straightforward, but there are several issues to consider. We interrogate both parts of this definition to argue that other verbal constructions have similar complexity and syntactic structure as a prototypical clause. Specifically, this section reviews constructions with coordination, nonfinite verbs, and small clauses, using examples[2] produced by low-advanced English L2 learners during a speaking task.

First, coordination is a basic way of increasing complexity (Norris & Ortega, 2009). In fact, Hunt's (1965) T-unit and Foster, Tonkyn and Wigglesworth (2000) Analysis of Speech unit (AS-unit) separated independent clauses,[3] due in part to the overuse of coordinating conjunctions. Coordination of prepositional phrases or noun phrases within a sentence adds complexity, but not clausal complexity because these phrases do not include a verb (Pallotti, 2015). Consequently, the coordination of independent clauses, prepositional phrases, or noun phrases, does not affect the coding of clauses.

The coordination of verbal elements, however, can complicate clause coding, despite the well-reasoned guidelines offered by earlier work in the field. Hunt (1965) defined a clause as a finite verb or coordinated verbs (not specifically verb phrases). According to Pallotti (2015), the coordination of verbs (e.g., *I sang and danced*) does not increase clause complexity because the verbs have a shared subject and seem to reflect a single event. Yet, an implied subject is possible in clauses (e.g., Foster et al., 2000; McManus et al., 2021), and coordinated verb phrases may reflect separate events. For instance, each coordinated verb phrase shown in (1) has its own complement, and they seem to be separate, sequential, events.

(1) *one night I took my father's car and* ***went to different city***

Admittedly, it can be difficult to determine whether the coordinated verb phrases reflect separate events/states. In (2) the coordinated verb phrases share a subject within a subordinate clause, which muddles whether '*feel upset*' reflects a separate event than '*was sad.*'

(2) … *when I was sad or* ***feel upset***[4]

The use of the disjunctive coordinating conjunction (i.e., '*or*'), however, suggests that these are separate events and thus separate clauses despite the shared subject within the adverbial clause. Besides, it would certainly complicate the clause coding for syntactic complexity if the semantics of the conjunction (e.g., '*or*' vs '*and*') affects the clause status of coordinated verb phrases.

[2] Any corrections, repetitions, and pausing information were removed from all examples except when that information was relevant to the analysis.

[3] There are no "compound sentences" when following Hunt (1965) guidelines for T-units or Foster et al.'s (2000) guidelines for AS-units (cf. Hwang et al., 2020).

[4] This L2 English example also includes a grammatical error; coding complications caused by grammatical errors will be discussed in a later section.

Moreover, given the clausal status of sentences without overt subjects in pro-drop languages (e.g., McManus et al., 2021), it seems reasonable that coordinated verb phrases have clause status when each verb has its own complement or adjunct phrase as in (1) and (2), regardless of whether we interpret the verb phrases as separate events. Indeed, some researchers (e.g., Menke, 2024; Vercellotti & Packer, 2016) included coordinated finite verb phrases in their clausal complexity measures, and Foster et al. (2000) went even further by suggesting coordinated verb phrases may be their own "independent" unit in some dialogues.

Now that we have established that overt subjects are not required in clauses, we can consider the requirement of conjugation, which is a defining feature of finite clauses. In English, the so-called "future tense" includes a modal and lexical verb, neither marked for tense or subject-verb agreement. In (3), the lexical verbs are unmarked as the coordinated verb phrases share a modal (i.e., '*would*').

(3) *she would be very patient and* ***comfort me***

The verb in the bolded unit retains the clausal feature of having a complement (i.e., *me*). In fact, the bolded unit in (3) is similar to the bolded unit in (2). In addition, (3) has the same structure as simple English imperatives, which are unequivocally considered clauses. It is unclear how '*comfort me*' uttered as an imperative could be considered a clause while '*comfort me*' as part of a coordinated verb phrase would not be considered equally complex. Accordingly, this type of coordinated verb phrase should be counted as a clause, too. Going just a bit further, in (4), the two verbs share a subject (i.e., '*she*'), modal (i.e., '*will*'), and complement (i.e., '*me*').

(4) *she will encourage and* ***accompany me****…*

The utterance in (4) may even represent a single event, leaning one to argue that (4) is a single clause. On the other hand, the structure of the bolded unit in (4) is syntactically equivalent to the bolded unit in (3), which has been shown to qualify a clause. Given the similarity of the language produced, a principled coding of clausal complexity includes the coordinated verb phrases of the types shown in (3) and (4) despite lacking an overt subject and tense morphology.

Second, nonfinite clauses also lack an overt subject and tense morphology. Since at least Hunt's (1965) definition of T-units, researchers (e.g., Menke, 2024; Vajjala & Meurers, 2012) have often limited clauses to *finite* clauses. In a summary of syntactic complexity indices, Kyle (2016: p. 25) stated "a clause is defined as a subject and a finite verb, though some studies … include clauses with non-finite verbs." The contrast in the quote reveals the default status of *finite* clauses (cf. Bulté & Roothooft, 2020; McManus et al., 2021; Vercellotti & Packer, 2016). Importantly, Foster et al. (2000) stated that a nonfinite verb with another element (i.e., complement or adjunct) reaches clause status. The production of another element creates additional complexity on par with finite clauses. Following Foster et al. (2000) coding guidelines, the bolded nonfinite clause in (5) would be coded as a separate clause because the nonfinite verb (i.e., *driving*) has a complement (i.e., *his car*).

(5) *there's someone* ***driving his car***

It should be noted that Pallotti (2015) suggested a slightly more demanding criterion, limiting clause status to nonfinite clauses when they denote a different action or state from the main clause. We saw, however, that determining whether the wording denotes separate events can be difficult to ascertain, particularly in an utterance with an existential '*there*,' as in (5). We argue that the clausal complexity comes from the learner's production of the additional syntactic element (i.e., a complement or adjunct), and so the bolded construction in (5) is a clause, like the coordinated verb phrases in Examples 1–4. Plus, relying only on the syntactic criterion of having its own complement, facilitates consistent coding.

After accepting that nonfinite clauses with another element reaches clause status, the complications continue because nonfinite clauses can also be coordinated. In (6), each nonfinite clause has an overt pronoun complement, but they share the infinitival '*to*'.

(6) *I hope to meet her and* ***have fun with her*** *again*

Considering that we have already accepted that modal auxiliaries are not strictly required in coordinated constructions such as in (3) and (4), it is reasonable to count the infinitive constructions in (6) as two separate (nonfinite) clauses because each has its own complement. A pattern emerges that a *clause* is either a finite verb or a nonfinite verb with a complement or adjunct.

Third, we should consider "small clauses" (Basilico, 2003). For instance, '*I found him fascinating*' is more complex, or at least more sophisticated, than a simple sentence because of the predication relation between "*him*" and the adjective phrase '*fascinating*' (Basilico, 2005). The bolded verbless small clause in (7) illustrates a similar relationship where '*them*' is described by the adjective phrase '*happy*' within a nonfinite clause (i.e., *to make them happy*).

(7) *the old people ask me to drink a coffee for them and to make* ***them happy***

Chen, Alexopoulou and Tsimpli (2021) stated that verbless small clauses are subordinate clauses, but they did not expound with a justification. In contrast, Kubo (1993) argued that these are syntactically noun phrases. As a matter of fact, verbless small clauses, like the bolded words in (7), syntactically resemble non-prototypical noun phrases such as '*someone new.*' Accordingly, without an overt verb, verbless "small clauses" fail to reach clause status.

In contrast, verbal small clauses have an overt verb, albeit one without verbal morphology. In (8), the verb '*open*' in the small clause has both a subject (i.e., '*them*') and a complement (i.e., '*the basket*').

(8) *in the picture number five we can see* ***them just open the basket***

Although Pallotti (2015) suggested that verbal small clauses do not typically qualify as a clause because the sentence/utterance depicts only one action/event, the bolded verbal small clause in (8) has both a verb and another syntactic element. Therefore, verbal small clauses have similar complexity to other clauses and should be coded as a subordinate clause.

In summary, we argue that coordinated verb phrases, nonfinite clauses, and verbal small clauses can reach the complexity of a clause with the learner's production of another element while verbless small clauses do not. Researchers can recognize these complex clauses as commensurate with finite subordinate clauses by coding them as subordinate clauses. Excluding constructions which are structurally similar to accepted clauses (e.g., imperatives) from the calculations of clauses requires a seemingly arbitrary division along a continuum of complex verbal constructions, which threatens the validity of the clause compleixty measure.

Coding complications caused by inaccuracies

The grammaticality of L2 language is part of the construct of accuracy, but certain errors can complicate measures of clausal complexity. For instance, the English predicate nominal construction may be produced by learners without a copula (e.g., Tode, 2007). As shown in (9), such utterances successfully describe a state because the copula complement comments on the subject.

(9) *it different from China*

If the coding system requires (or depends upon) the production of a verb, utterances like (9) may be inadvertently excluded from the clause calculations. The failure to recognize these clauses would skew measures of clausal complexity, particularly for certain populations, such as learners whose first language does not require the copula in all contexts (e.g., Arabic, Mandarin) and would misrepresent complexity scores of learners from different language backgrounds, coloring comparisons within and across research studies.

Furthermore, verb morphology errors are common in L2 production. For instance, for English verbs with aspect, tense is found on the required auxiliary. In (10), it seems as though the learner is attempting to form progressive aspect, but the auxiliary was not produced.

(10) *so we* ***enjoying*** *the very special very traditional Russian food there*

Without the required auxiliary, the verb with its *-ing* morphology may be coded as a participle or gerund, rather than a finite verb. Notably, regardless of the interpretation (nonfinite or finite), it would be coded as a clause following the broader definition of clause presented here.

Unfortunately, most research articles (e.g., Kuiken & Vedder, 2011; Leonard & Shea, 2017) do not acknowledge any complications of coding clausal complexity in learner data (cf. Vercellotti, 2017). When investigating the reliability of automatic parsers on advanced L2 written data, Lu (2010) concluded that most inaccuracies did not cause coding errors for the chosen measures. With a corpus of texts written by learners across a range of proficiencies, however, Huang, Murakami, Alexopoulou and Korhonen (2018), found that some verb morphology inaccuracies did cause coding errors by the automated parsers. To address these issues that may thwart clause identification, researchers must clearly articulate any adjustments made when coding nonstandard forms (Lintunen & Mäkilä, 2014), pre-process the data to improve parser success (e.g., Kim & Lu, 2024), or review part-of-speech tags before the clausal complexity calculations.

Complications of speech data

L2 speaking generally requires more cognitive resources than writing (Hwang, Jung & Kim, 2020), and there is pressure to maintain the speech stream (Kuiken & Vedder, 2011). As a result, speech data tend to have more language errors than written texts, increasing the coding complications discussed in the previous section. Additionally, speech data have more "noise" such as repetitions and false starts (Kim & Lu, 2024). Foster et al. (2000) defined the AS-unit with consideration of many complications of segmenting spoken data, including the frequent use of transition words. Despite their guidance, complications still arise from the extensive linking of ideas, which is more common in speech, as shown in (11).

(11) *if I work hard* ***or*** *if I think hard* ***or*** *if I do stuff and work more I'll get more money*

The utterance begins with three coordinated subordinate clauses beginning with '*if*', and it is difficult to decide if the speaker intended a list of events or if the speaker was trying various options to express the same idea (i.e., self-correcting). Certainly, grammar and lexical corrections are common in speech (Vercellotti & McCormick, 2018). If this utterance had been written with the opportunity to be edited, perhaps the first two clauses would have been replaced by the third. Given that each clause is fully grammatical and there is no clear lexical substitution, each could be intended, and therefore, plausibly coded as separate clauses. The next example illustrates an even bigger challenge. The utterance in (11) was followed by the utterance in (12), which includes the non-lexical filled pause *uh* and silent pauses marked with (.).

(12) *and that* (.) *that which* (.) *uh* (.) *make it a good* (.) *job for me* (.)

Even with the additional pausing (and intonation) information available in speech data, it is unclear whether the speaker is self-correcting or adding clauses. When coding clauses in data with superfluous connector words, corrections, and errors, such as the utterances shown in (11) and (12), researchers should acknowledge any systematic complications and resolutions in the methodology section.

The final complication to (briefly) address is the finite clause type which takes an entire clause as its linguistic complement (e.g., direct object), as shown in (13).

(13) *uh* (.) ***I think***[5] (.) uh (.) *we should* (.) *uh* (.) *uh* (.) *do some protection to:* (.) *protect our* (.) *world* (.)

These clauses are likely to be matrix verbs when in initial position and with the complementizer '*that*' (Dehé & Wichmann, 2010). However, '*I think*' and other "minor clauses" (Fung & Carter, 2007) can express epistemic stance (Brown, Fernández & Huensch, 2023), similar to adverbs such as '*maybe*' or '*possibly*'. Dehé and Wichmann (2010) have convincingly argued that when used in medial or final position, '*I think*" functions as a (grammaticalized) adverb. Hence, the '*I think*" in (14) is emblematic of its use as an epistemic hedge.

(14) *and* (.) *you know you can because* (.) *because every team* ***I think*** *should have a* (.) *good leader* (.) *like that* (.)

Certainly, it is difficult to differentiate between the use of '*I think*' as an epistemic hedge and a matrix verb of cognition. Given the frequency of this verb (Vercellotti, Juffs & Naismith, 2021), particularly with topic prompts which ask for an opinion, researchers should recognize how these clauses inflate the number of clauses while lowering the mean length of clause. Given their syntactic structure, such clauses are clauses; however, some studies (e.g., Dehé & Wichmann, 2010; Vercellotti & Packer, 2016) have coded these clauses separately from matrix clauses, which allows them to be tallied and analyzed separately if warranted.

Conclusion

This commentary has demonstrated that some verbal constructions beyond the prototypical clause demonstrate clausal properties and comparable syntactic complexity to finite clauses. Accordingly, the term *clause* should be expanded to include verb phrases and nonfinite clauses with another syntactic element. Including verb phrases with a complement and/or adjunct as a separate clause is justified given the structural similarity (and thus similar level of complexity) with imperatives, which are accepted as clauses. The explicit inclusion of verb phrases with a complement/adjunct in the clause coding allows for consistency across studies, including research with certain L2 English learners (e.g., those with a pro-drop language background) who might fail to produce a subject as well as research with L2 learners of pro-drop languages.

Additionally, the field should recognize nonfinite clauses as subordinate clauses, finally adopting the suggestion that Foster et al. (2000) made nearly a quarter of a century ago. Giving clause status to nonfinite clauses may also mitigate confusion caused by missing or superfluous verbal morphology in learner data. Likewise, a clause which lacks a copula but clearly describes a state should be considered a clause, which is often grammatically acceptable cross-linguistically. In summary, the definition of *clause* in L2 research should include:

- verb phrases with a complement and/or adjunct, such as (1) and (2)
  - including verb phrases with a shared modal, such as (3) and (4)
- nonfinite clauses with a complement and/or adjunct, such as (5)
  - including nonfinite clauses with a shared infinitival '*to*,' such as (6)
  - including verbal small clauses, such as (8)
- copula constructions without the copula, such as (9)

This definition, supported by previous work in the field, is syntactic, and each of the above would be subordinate clauses in the complexity calculations. Verbless small clauses, such as (7), however, would be excluded from the definition of clause. To supplement the examples discussed above, Appendix A shows a coded transcript of a speech produced by a low-advanced English learner. The coding advocated here results in 23 clauses in thirteen AS-units (1.77 clauses/AS-unit), reflecting that many of the AS-units were multi-clausal.

Further, having this default definition of *clause* replaces a potentially lengthy description of what verbal constructions were included in the coding scheme. If a researcher decides to operationalize *clause* more narrowly, which may be appropriate either theoretically or for practicality, they can report the inclusion criteria briefly, such as "finite clauses with an overt subject." As a point of comparison, if the transcript in Appendix A were coded following a narrower definition, only 15 (bolded) clauses would be coded, and the same speech would seem to have few multi-clausal AS-units (1.15 clauses/AS-unit).

Although the examples described here reflect issues that emerged from just one data set, it is likely that all L2 data have some of

[5] In addition, '*I think*' may sometimes serve as a verbal filler and a repair strategy in speech (Brown et al, 2023; Dehé & Wichmann, 2010). In (13) the '*I think*' may be used for fluency as it begins with a non-lexical filler, a pause, and another non-lexical filler before offering an opinion.

these complex (but often excluded) clauses and some difficult-to-code clauses. The field's acceptance of this broader definition of *clause* will facilitate the interpretation of research findings on a range of syntactic complexity measures, including the commonly used measures of mean clause length and clauses per sentential unit. Clarity in the research methodology enables readers to better interpret findings. A clear, shared, definition of *clause* also allows better comparison of results across studies. We ask the field to use this justified, broader definition of clause, and we invite the field to continue discussing coding complications found in L2 data.

CRediT authorship contribution statement

MaryLou Vercellotti: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Data curation, Conceptualization. **Sean Hall:** Writing – review & editing, Data curation.

Declaration of competing interest

None

Acknowledgements

Sincere thanks to the journal's anonymous reviewers whose thoughtful feedback greatly strengthened this paper. Thanks also to Delphine Rapenne who asked good clause coding questions. Open Access publication of this commentary was supported by Aspire Published Scholarship Support through Ball State University, Muncie, Indiana.

Appendix A[6]

well my favorite place in general it's being in desert especially at night **[^ci]** (.)
but &-ah here in United_State(s) I [/] I have heard about many places very very interesting places like (.) &-uh (.) &-uh (.) Canyon Valley and the Smokey Mountain and a lot a lot of places here **[^ci]** (.)
but &-yeah (.) &-uh (.) going back to my favorite place [^cnf] which is desert especially Libyan desert **[^cr]** we have a lot of (.) &-uh (.) lakes (.) in the deserts and (.) &-uh (.) palms (.) &-uh (.) a lot of palms tree **[^ci]** (.)
so &-yeah (.) &ga going (.) &-uh (.) over there in the (.) desert at &-ah night [^cnf] and try [^cvp] to hunting some animals [^cnf] (.) &-uh (.) it's will be (.) more than great to me **[^ci]** (.)
and spending a night [^cnf] (.) &-uh (.) it will be cold **[^ci]** (.)
and (.) &-ah (.) around (.) &-uh (.) the fire that we'll probably will make **[^ci]** (.) and (.) &-uh (.) telling our travel stories [^cnf] (.) &-uh (.) &-yeah (.) &-yeah I've nothing more **[^ci]** we can say about desert **[^cr]**
but you can (.) &-uh (.) try it **[^ci]** (.)
can't describe it to you **[^ci]**
it's one of the hardest things to describe **[^ci]** (.)
it will be warm **[^ci]** (.) to be &-ah surrounding by friends (.) inside the &ser desert at night [^cnf] (.)
I: can't (.) explain **[^ci]** or express my feeling right now [^cvp] (.)
I'm kind of missing (.) this feeling (.) &-yeah **[^ci]** (.)

References

Ahmadian, M. J., Tavakoli, M., & Dastjerdi, H. V. (2015). The combined effects of online planning and task structure on complexity, accuracy and fluency of L2 speech. *The Language Learning Journal, 43*(1), 41–56.
Basilico, D. (2003). The topic of small clauses. *Linguistic Inquiry, 34*(1), 1–35.
Brown, D., Fernández, J., & Huensch, A. (2023). The role of pragmatic markers in perceptions of L2 fluency in dialogue. *System, 119*. https://doi.org/10.1016/j.system.2023.103157
Bulté, B., & Roothooft, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System, 90*, Article 102246. https://doi.org/10.1016/j.system.2020.102246
Chen, X., Alexopoulou, T., & Tsimpli, I. (2021). Automatic extraction of subordinate clauses and its application in second language acquisition research. *Behavior Research Methods, 53*, 803–817.
Dehé, N., & Wichmann, An. (2010). The multifunctionality of epistemic parentheticals in discourse: Prosodic cues to the semantic-pragmatic boundary. *Functions of Language, 17*(1), 1–28.
De Jong, N., & Vercellotti, M. L. (2016). Similar prompts may not be similar in the performance they elicit: Examining fluency, complexity, accuracy, and lexis in narratives from five picture prompts. *Language Teaching Research, 20*(3), 387–404.

[6] Transcriptions follow the TalkBank format (MacWhinney, 2000) with pauses (.), repetitions [/], and corrections [//], filled pauses (&-uh), and partial words (&gr) indicated. AS-units are presented on separate lines. Clauses are marked with [^c] with independent [^ci], relative [^cr], nonfinite [^cnf], and coordinated verb phrases [^cvp] specified. Subordinate adverbial clauses and "minor clauses" would be included in both the broad and narrow coding, but none were produced in this speech.

el Majidi, A., de Graaff, R., & Janssen, D. (2021). Debate as a pedagogical tool for developing speaking skills in second language education. *Language Teaching Research*. https://doi.org/10.1177/13621688211050619
Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*, 354–375.
Fung, L., & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogical settings. *Applied Linguistics, 28*(3), 410–439.
Huang, Y., Murakami, A., Alexopoulou, T., & Korhonen, A. (2018). Dependency parsing of learner English. *International Journal of Corpus Linguistics, 23*(1), 28–54.
Hunt, K. W. (1965). A synopsis of clause-to-sentence length factors. *The English Journal, 54*(4), 300+305. -309.
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.
Kim, M., & Lu, X. (2024). L2 English speaking syntactic complexity: Data preprocessing issues, reliability of automated analysis, and the effects of proficiency, L1 background, and topic. *The Modern Language Journal*. https://doi.org/10.1111/modl.12907
Kubo, M. (1993). Are subject small clauses really small clauses? *MIT Working Papers in Psycholinguistics, 3*, 93–115.
Kuiken, F., & Vedder, I. (2011). Task complexity and linguistic performance in L2 writing and speaking: The effect of mode. In P. Robinson's (Ed.), *Second language task complexity* (pp. 91–104). John Benjamins.
Kyle, K. (2016). *Unpublished doctoral dissertation*. Georgia State University.
Leonard, K. R., & Shea, C. E. (2017). L2 speaking development during study abroad: Fluency, accuracy, complexity, and underlying cognitive factors. *The Modern Language Journal, 101*(1), 179–193.
Lintunen, P., & Mäkilä, M. (2014). Measuring syntactic complexity in spoken and written learner language: Comparing the incomparable? *Research in Language, 12*(4), 377–399.
MacWhinney, B. (2000). *The childes project: Tools for analyzing talk* (3rd ed). Lawrence Erlbaum Associates.
McManus, K., Mitchell, R., & Tracy-Ventura, N. (2021). A longitudinal study of advanced learners' linguistic development before, during, and after study abroad. *Applied Linguistics, 42*(1), 136–163.
Menke, M. R. (2024). Beyond accuracy: The development of syntactic complexity in middle school Spanish immersion. *Foreign Language Annals*. https://doi.org/10.1111/flan.12746
Norris, J. M., & Ortega, L. (2009). Towards an organic approach to investigating CAF in instructed SLA: The case of complexity. *Applied Linguistics, 30*(4), 555–578.
Pallotti, G. (2015). A simple view of linguistic complexity. *Second Language Research, 31*(1), 117–134.
Skehan, P., Bui, G., Wang, Z., & Shum, S. (2024). Re-examining accuracy measures in second language task-based spoken performance. *Research Methods in Applied Linguistics, 3*. https://doi.org/10.1016/j.rmal.2024.100098
Tode, T. (2007). Durability problems with explicit instruction in an EFL context: The learning of the English copula be before and after the introduction of the auxiliary be. *Language Teaching Research, 11*(1), 11–30.
Vajjala, S., & Meurers, D. (2012). On improving the accuracy of readability classification using insights from second language acquisition. In *The 7th Workshop on the Innovative Use of NLP for Building Educational Applications* (pp. 163–173).
Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics, 38*(1), 90–111.
Vercellotti, M. L. (2018). Finding variation: Assessing the development of syntactic complexity in ESL speech. *International Journal of Applied Linguistics, 19*(2), 233–247. https://doi.org/10.1111/ijal.12225
Vercellotti, M. L., Juffs, A., & Naismith, B. (2021). Multiword sequences in English language learners' speech: The relationship between trigrams and lexical variety across development. *System, 98*, Article 102494. https://doi.org/10.1016/j.system.2021.102494
Vercellotti, M. L., & McCormick, D. E. (2018). Self-correction profiles of L2 English Learners: A longitudinal multiple-case study. *TESL-EJ, 22*(3). https://tesl-ej.org/wordpress/issues/volume22/ej87/ej87a3/.
Vercellotti, M. L., & Packer, J. (2016). Shifting structural complexity: The production of clause types in speeches given by English for academic purposes students. *Journal of English for Academic Purposes, 22*, 179–190.