

Emphatic Accentuation and Discourse Intention in L2 English Speech

Production: Evidence from a Picture-Description Task

Hong Zou

School of English Language and Culture, South China Business College,
Guangdong University of Foreign Studies

Author Note

The author declares no conflicts of interest to disclose. Correspondence concerning this article should be addressed to Hong Zou, School of English Language and Culture, South China Business College, Guangdong University of Foreign Studies, Guangzhou, Guangdong, China. Email: phylliszou@126.com

Abstract

Cognitive research has shown that the prosodic realization of emphatic accent reflects a speaker's discourse intention. This study compares two native speakers of British English and eight Chinese learners in a picture-description task, focusing on IP-focus alignment, prenuclear-nuclear accent coordination, and the re-accenting of given information. Native speakers organized speech in a listener-oriented manner: IPs typically contained a nuclear accent for closure and, when needed, a prenuclear B-accent to establish contrast or a discourse frame. Learners, by contrast, produced denser, less hierarchically differentiated accent patterns, often placing several strong accents within one IP and using given-information accents mainly for signposting or repetition. These differences suggest distinct communicative orientations: native speakers designed utterances for a potential listener, whereas learners prioritized task completion and propositional coverage. The findings support integrating discourse awareness and prosody-information-structure alignment into L2 speaking instruction and assessment, with the three dimensions serving as operational indicators of communicative orientation.

Keywords: discourse intention, emphatic accentuation, B-accent, information structure, L2 English speech production

Language is not only a vehicle for transmitting information but also a resource for expressing communicative intentions. In spoken English, accentuation is a central prosodic device for making grammatical and semantic distinctions while also enabling speakers to make

Received 20 July 2026; revised 25 August 2026; accepted 10 September 2026; available online 25 September 2026; version of record 30 September 2026.

Citation: Zou, H. (2026). Emphatic Accentuation and Discourse Intention in L2 English Speech Production: Evidence from a Picture-Description Task. *Journal of Language*, 2(3), 165–196.
<https://doi.org/10.64699/26FJHV6418>

particular meanings salient. The distribution of accents in second-language (L2) speech can therefore provide direct evidence of how learners conceptualize and organize what they intend to say.

Previous research on L2 English prosody has largely concentrated on systematic differences between learners and native speakers, including differences in prominence, phrasing, fluency, and information-structural marking (Hahn, 2004; Jun, 2006; Trofimovich & Baker, 2006; Kang, 2010; Gut & Pillai, 2014). Much less attention has been paid to one possible source of these differences: the discourse intention that guides speech planning. The present study compares accentuation in picture descriptions produced by native speakers and Chinese learners of English. It shows that the native speakers remained listener-oriented even when no actual listener was present, whereas the learners displayed a strong orientation toward completing the task. This contrast was reflected in the native speakers' use of emphatic accentuation to guide listeners through the narrative, whereas the learners used it primarily to verbalize the content depicted in the pictures.

The study therefore treats accent placement not simply as a matter of phonetic accuracy but as the observable outcome of communicative planning. By examining how speakers package information within intonational phrases, coordinate prenuclear and nuclear accents, and re-accent given information, the study seeks to connect prosodic form with discourse intention and to identify pedagogically useful indicators of listener-oriented speech.

1. Discourse Intention, Focus, and Accentuation

Levelt's (1989) model of speech production proposes that speaking is guided by the speaker's intention. Production proceeds through conceptualization, linguistic formulation and

articulation, together with self-monitoring. Emphatic accentuation can be understood as the phonological projection of concepts selected for prominence under the guidance of discourse intention. In the present study, the term covers both narrow-focus accents in the tradition of Halliday and B-accent in Büring's (2003) contrastive-topic-focus (CT-F) structure.

Within information structure, narrow-focus accentuation marks new or contrastive information (Halliday, 1967). Such information often corresponds to a set of semantic alternatives or a pragmatic update (Rooth, 1992; Schwarzschild, 1999). Building on the Nuclear Stress Rule and principles of intonational phrasing, focus accentuation has also been treated as an interface marker between syntax and semantics, with a focus domain projected from the right edge of the syntactic structure (Chomsky & Halle, 1968; Kahnemuyipour, 2009). In English, the focus domain commonly aligns with an intonational phrase (IP), and the focus accent is frequently realized as the nuclear accent that supplies semantic and prosodic closure (Selkirk, 1995; Ladd, 2008; Gussenhoven, 2004). Narrow focus is typically realized by highly salient pitch accents such as H*, L+H*, and H*+L (Pierrehumbert, 1980).

Büring (2003) extends the analysis beyond focus by proposing B-accent on contrastive topics and other nonfocal material. In narrative discourse, speakers may accent a topic or another nonfocal element to separate information units, establish a dimension of contrast, and prepare for the resolution or closure delivered by the focus accent. The interaction between CT-F structure and accentuation can thus be represented as a discourse tree in which prenuclear and nuclear accents perform complementary functions.

2. Research Questions

The importance of accentuation for discourse comprehension has been repeatedly

demonstrated in perception and processing studies. Birch and Clifton (1995, 2002), for example, manipulated accent placement in question–answer pairs and found that responses accented on the focused constituent were judged more appropriate and processed more quickly. Hahn (2004) reported a similar pattern: when a lecture excerpt produced by an international teaching assistant was presented in versions with primary stress on the focus, on a nonfocus, or with no clear primary stress, native listeners recalled the most information from the focus-aligned version. Eye-tracking research by Dahan, Tanenhaus, and Chambers (2002) likewise showed that an accented noun is more readily interpreted as introducing a new, nonanaphoric, or contrastive referent and consequently attracts additional attention.

Two conclusions follow from this work. First, emphatic accentuation marks the information that the speaker intends to foreground and simultaneously provides the listener with a perceptual cue to that information. Second, when the distribution of accents conflicts with listeners’ expectations, processing becomes more difficult and may lead to misinterpretation. Accentuation problems in Chinese learners’ English therefore concern not only pronunciation accuracy but also communicative intelligibility.

Under Levelt’s (1989) model, the formal properties of an utterance are first shaped by discourse intention and are subsequently monitored against the speaker’s own linguistic experience. If the speaker does not conceptualize the utterance as being addressed to a listener, the resulting accentual pattern may be poorly adapted to comprehension. To examine this hypothesis, the present study used a picture-description task to address two questions:

(1) What marked differences occur between native speakers and Chinese L2 learners in the use of emphatic accentuation in English speech production?

(2) To what extent do these differences reflect different discourse intentions?

3. Method

3.1 Participants

The participants were recruited from a university in Guangzhou, China. The learner group consisted of eight volunteer English majors in their third or fourth year of study. The comparison group comprised two British English-speaking instructors employed at the same institution, one woman and one man. Table 1 summarizes the participants' background information.

Table 1

Participant Information

Participant	Sex	Age	Everyday language use	TEM-4 score
BPW	F	21	Mandarin 90%; Hakka 10%	78
LZX	M	22	Mandarin 70%; Cantonese 30%	67
YAQ	F	22	Mandarin 80%; Nanxiong dialect 20%	54
FJC	M	21	Mandarin 90%; Teochew 10%	65
HZY	F	21	Mandarin 90%; Hakka 10%	58
GYG	F	21	Mandarin 70%; Cantonese 30%	56
HXT	F	20	Mandarin 60%; Cantonese 40%	60
ZYB	M	20	Mandarin 50%; Teochew 20%; Hakka 30%	75
A	F	34	English 90%; Urdu 10%	—
N	M	48	English 98%; Korean 2%	—

The learner participants came from different parts of Guangdong Province and used both Mandarin and regional varieties in daily life, although Mandarin was the dominant language for all eight. The group included three men and five women. Their mean Test for

English Majors Band 4 (TEM-4) score was 64.1 (SD = 8.24), indicating some variation but no extreme dispersion in general English proficiency.

3.2 Materials and Procedure

All ten participants were recorded in a microteaching laboratory at the university using professional recording equipment. They received the task materials 20 minutes before recording. During preparation, they were permitted to use a mobile phone to check word meanings and phonemic transcriptions, but not to use audio pronunciation functions. Pauses, repetitions, and other performance disfluencies were permitted during recording.

The task consisted of five pictures illustrating the fable “The North Wind and the Sun” (Figure 1). All participants reported that they were familiar with the story. The picture sequence provided a relatively constrained discourse context while still allowing speakers to decide which events, participants, and background changes should be made prominent. The distribution and function of emphatic accents in the resulting narratives were therefore used as evidence of speakers’ discourse planning and communicative orientation.

Figure 1

Pictures Used in the “North Wind and the Sun” Description Task



4. Analysis and Findings

According to Levelt's (1989) model, a listener-oriented description of this picture sequence should, at the level of macroplanning, elaborate the overall communicative goal into a series of subgoals and select and order the information needed to convey the central issue of which character is stronger. At the level of microplanning, the speaker should assign the information an appropriate propositional form and informational perspective, including topic–focus structure, so as to guide the listener's attention.

The two native speakers systematically implemented this listener orientation through their organization of accents. Although their narrative styles are different, both produced patterns that were grounded in the visual material and designed for a listener. The learners' accentuation, by contrast, was less listener-oriented and more closely resembled task-completion production. There are three distinct characteristics: learners often placed several strong accents within a single IP, creating oversized information packages; their accents tended to function independently rather than through coordinated focus and B-accent relations; and their re-accenting of given information served a narrower range of discourse functions.

4.1 Information Packaging Within IP

In Gussenhoven's (1983, 1984, 2004) account, the focus domain is delimited by the nuclear accent and its projection and normally aligns with an IP boundary. This focus-domain–IP correspondence allows prosodic segmentation—through boundaries, closure, and pitch reset—to coincide with information-structural integration, such as new versus given information and contrast versus resolution. Experimental work suggests that stacking multiple strong accents within a single IP increases the demands of attentional allocation and structural

recovery (Birch & Clifton, 1995; Dahan et al., 2002; Kjelgaard & Speer, 1999).

This correspondence was pervasive in the native-speaker data. Speaker A produced an average of approximately 4.6 words per IP and Speaker N approximately 5.1 words. An IP generally contained only one or two emphatic accents: the nuclear accent functioned as the focus accent, sometimes accompanied by a prenuclear B-accent. Native-speaker IP boundaries were marked by clear pauses of at least approximately 0.5s, preboundary lengthening, and postboundary pitch reset. These cues provided the listener with clear signals of segmentation and closure, consistent with previous research (Wightman et al., 1992; Krivokapić, 2007).

The learners, in contrast, produced an average of approximately three emphatic accents per IP. Their IPs were strongly dependent on grammatical structure, and for some participants the number of IPs closely matched the number of grammatical sentences. Within an IP, short pauses often divided the utterance into smaller units, most of which contained one or two accents. These smaller units lacked full closure and pitch reset and resembled intermediate phrases (ip) in the ToBI framework (Beckman & Ayers Elam, 1997; Jun, 2006). When several ips were juxtaposed within a single IP, the pitch pattern often alternated between H* and !H*. In English, an H*–!H* sequence commonly signals a relation between focus and associated secondary information (Ladd, 2008). When multiple accents are placed side by side without clear hierarchy, however, the listener cannot readily determine where intonation closes or where the discourse advances and therefore cannot recover a coherent information unit.

Under these conditions, listeners must process the utterance along two partially independent tracks. The one infers the narrative focus from accent placement while the other potential boundaries from pauses and grammatical structure. Because the two tracks lack a

stable point of integration, the utterance offers few reliable processing anchors and imposes a greater cognitive burden. Example (1), produced by FJC, illustrates this pattern.

- (1) They SAW(H*) a TRAVELLER*(H*) walking along the ROAD(H*+L)/ and
DECIDED(H*+L) to USE(H*) him as a test/to test the subject to see/ WHO*
(H*) could make the TRAVELLER(H*) remove his coat FIRST(H*+L).¹

The utterance lasted approximately 10.19 s and contained three pauses, the longest of which was about 0.3 s, dividing the IP into four ips. Pitch fell before the pauses, but there was no preboundary lengthening and no substantial postpause pitch reset². Three of the ips contained two or three emphatic accents. Because the accents were densely distributed and boundary cues were weak, a listener would have difficulty identifying transitions between stages of the narrative, thereby reducing the efficiency of online comprehension.

A native speaker conveying the same information would be more likely to distribute it across three or four IPs, each with a lower accent density and a boundary that provides a stable point of closure for the focus. The important issue is not whether the utterance is grammatically correct or whether its propositional content can ultimately be recovered. Rather, a mismatch between the information package and the IP increases the processing burden on a potential

¹ In annotating the intonation of emphatic accents, effects attributable to participants' habitual phonetic realizations of particular segments were excluded. For example, the final syllable of *traveller* was consistently produced in an extra-low pitch range throughout the picture-description task. This low-pitched portion was therefore disregarded, and the accent was labeled on the basis of the stressed syllable, resulting in an H* annotation. Capitalization marks emphatic accentuation, and pitch-accent labels are provided in parentheses. H* denotes a high pitch accent, while H*+L denotes a high accent followed by a low trailing tone. A single slash (/) marks an intermediate-phrase (ip) boundary. An asterisk following a word, outside the parentheses, indicates that the word carries a relatively stronger accent than the other accented words within the same ip. The same notation is used in the subsequent examples.

² In all relevant cases, the fall was realized as an abrupt drop in the second half of the final syllable. The F0 onset of the first syllable after the pause was approximately equal to the first half of the final syllable before the pause, indicating little or no reset.

listener. Listener-oriented planning distributes prominent information across IPs in a relatively even and interpretable manner; the learner pattern suggests that the central goal was to complete the description rather than to design it for a listener.

4.2 Coordination Between Focus Accents and B-accents

Two emphatic accents within a single English IP are not unusual. The rightmost accent is the nuclear accent and provides closure for the information unit, whereas accents to its left are collectively termed prenuclear accents. A prenuclear accent is not necessarily an independent focus. Instead, it supplies a cue that works together with the nuclear accent to encode information structure (Selkirk, 1995; Ladd, 2008; Gussenhoven, 2004; Büring, 2003). This complementary coordination creates what may be called a prenuclear–nuclear “leverage effect”: the prenuclear accent establishes a starting point or dimension, such as a contrast set, scalar relation, or topic shift, and the nuclear accent subsequently delivers a decision or closure. The two accents support one another while differing in prominence and discourse function (Pierrehumbert & Hirschberg, 1990; Büring, 2003; Ito & Speer, 2008; Watson et al., 2008).

Online-processing studies show that a prenuclear accent in a CT–F structure can proactively narrow the search space and direct visual attention toward an appropriate contrast set (Ito & Speer, 2008; Watson et al., 2008). Perception studies likewise demonstrate that well-formed accentual configurations reduce the burden of referential disambiguation and structural recovery, whereas multiple strong accents without hierarchical differentiation interfere with comprehension (Birch & Clifton, 1995; Dahan et al., 2002).

In the two native-speaker narratives, prenuclear accents accounted for 47% of all accents for Speaker A and 38% for Speaker N, indicating that double-accent IPs were common.

There are two relations of double-accent. In the dominant pattern, the nuclear accent was the focus accent and the prenuclear accent was a B-accent. This relation accounted for 69% of A's double-accent IPs and 72% of N's. The nuclear accent provided cumulative right-edge closure and was more prominent than the prenuclear accent. In this case, the prenuclear accent served three functions. First, it established a scalar dimension, as in SO HOT, where L+H* on SO triggered a scalar or contrastive inference. Second, it linked topic and predication, as in A's "The WEATHER looks NEUTRAL," where WEATHER established a discourse anchor and the nuclear accent on NEUTRAL supplied a concluding state. Third, it managed affect or discourse transition, as in N's "OH, that's ME," where a prominent OH signaled a change of cognitive state and ME resolved the self-reference (Heritage, 1984).

The second relation was more marked: the prenuclear accent functioned as the focus accent and the nuclear accent as a B-accent. This pattern occurred in 5 of A's 16 double-accent IPs and 7 of N's 25. It was typically associated with strong contrast or a special pragmatic effect. N, for example, produced "It was SO HOT" twice. On the first occurrence, HOT bore the main accent, conforming to the default rightward pattern. On the second occurrence, SO became the main accent with contrastive L+H*, while the given HOT functioned as a supporting B-accent. This configuration anticipated the subsequent scalar comparison in "He TOO couldn't bear it" and shows that the leverage relation can extend across sentences, with a dimension established in advance and resolved later in the discourse.

Among the learners, prenuclear accents accounted on average for 59% of all emphatic accents, ranging from 10% for BPW to 74% for the highest individual value. When calculated relative to ips, the proportion was approximately 33%. The resulting pattern of ips nested within

larger IPs appears to arise from two competing tendencies: learners aligned IPs with grammatical units, yet also attempted to accent a large number of conceptual elements. Smaller ips were therefore embedded inside grammatically defined IPs to create alternating strong–weak rhythms and to project multiple concepts onto the phonological form. This production design prioritized the expression of content but paid limited attention to how a listener might parse it.

(2) The SUN(H*) shone WARMLY*(H*+L) and BRIGHTLY(H*+!H) on the traveller.

Example (2), produced by LZXX, contains three emphatic accents in one IP. Their mean F0 values were 173 Hz, 126 Hz, and 127 Hz, respectively. The prenuclear accent on SUN was the strongest and can therefore be interpreted as the focus accent, whereas the nuclear accent on BRIGHTLY differed little in strength from the other prenuclear accent on WARMLY. The falling movement on the nuclear accent did not reach a sufficiently low endpoint to provide clear prosodic closure. In contextual terms, SUN was given information, yet its strong prenuclear accent did not produce a marked contrastive or intensifying effect. The three accents were thus accumulated without forming a functional hierarchy.

From such an IP, the listener receives little guidance regarding which information is central, which element establishes a contrastive dimension, and where the unit closes. Audience-design research predicts that speech style and prosodic resources are adjusted to the speaker’s model of the addressee (Bell, 1984; Clark, 1996). English intonational research further suggests that prenuclear accents establish dimensions or contrasts and nuclear accents close or resolve them. Their pairing and hierarchy are therefore important mechanisms through which intention is externalized for a listener (Pierrehumbert & Hirschberg, 1990; Ladd, 2008;

Gussenhoven, 2004).

Without a listener-oriented design, the problem is not simply insufficient stress, but a breakdown of accentual organization. Multiple strong accents may compete without a clear main point, or the main accent may fail to connect with its supporting accent. In such cases, the leverage relation no longer functions.

4.3 Re-accenting Given Information in Narrative Discourse

Listener-oriented speech must accommodate limits on processing and the principle of uniform information density (UID). Discourse should maintain a reasonably stable information rate, avoiding both excessively dense stretches that overload processing and excessively sparse stretches that reduce efficiency (Aylett & Turk, 2004; Levy & Jaeger, 2007; Jaeger, 2010). Re-accenting some given information can therefore be necessary. At points of locally low information density—such as transitions, turn shifts, and positions before or after the announcement of a result—speakers may prosodically reinforce given material to smooth the information flow, mark discourse structure, or establish a contrastive domain (Aylett & Turk, 2004; Bell et al., 2009; Wagner & Watson, 2010). Such re-accenting supports intelligibility and organization without adding new propositional content.

In the present data, emphatic accents in both groups fell predominantly on new information (at least 75%); for the two learners with the fewest accents, more than 90% of accents marked new information. Nevertheless, the two groups differed substantially in the discourse functions assigned to accents on given information.

Drawing on previous research, the present study classified the functions of re-accented given information into four categories: signposting, contrast, intensification, and repetition.

Signposting accents reactivate a referent, mark a discourse turn or scope, establish a contrastive domain or scale, or signal a reversal in semantic direction (Terken & Hirschberg, 1994; Baumann & Riester, 2012). Contrastive accents mark either a contrasted entity or the outcome of a contrast and are usually focus-like (Ladd, 2008). When contrast and signposting co-occurred at a transition or point of closure, contrast was coded as the primary function. Intensification was coded when re-accenting increased the perceived strength, stance, or threshold associated with an existing cue (Eady et al., 1986; Sluijter & van Heuven, 1996). Repetition referred to unnecessary re-presentation of given content that could be replaced by an unaccented form or a more economical expression without loss of meaning.

Each re-accented item was assigned one primary function and, where appropriate, one secondary function. Primary functions were used for the main descriptive statistics and inferences. Secondary functions were retained to capture functional overlap and to support a more fine-grained qualitative interpretation.

Table 2

Primary Functions of Emphatic Accents on Given Information

Group	Signposting	Contrast	Intensification	Repetition	Total
Native speakers	12 (48%)	4 (16%)	9 (36%)	0	25
Learners	40 (57%)	11 (16%)	4 (6%)	15 (21%)	70

As Table 2 shows, three principal differences emerged between the groups, each consistent with listener-oriented design in the native-speaker data and task-completion production in the learner data.

4.3.1 Primary Functions

First, signposting was the most frequent function in both groups, but its proportion was higher among learners and its hierarchical organization was less clear. HXT, for example, used two identically accented instances of FINALLY in successive stretches, one marking local closure and the other global closure, a pattern that could mislead a listener (Clark & Haviland, 1977). Similarly, three of ZYB's six re-accented given items fell on TRAVELLER and served the same signposting function of maintaining the affected participant, generating excessive redundancy.

The native speakers instead embedded signposts in chains of discourse progression consistent with a CT–F structure (Büring, 2003). N re-accented the given referents SUN and WIND at turn-initial, transitional, and closing points—"met first with the WIND," "it was the SUN'S turn," and "in the end, the SUN proved ..."—and coordinated these signposts with intensifying or contrastive expressions such as HARDER and SO HOT. The resulting sequence was signpost → contrast/intensification → resolution.

Second, the groups differed sharply in intensification and repetition. Only 6% of the learners' re-accented given items had intensification as their primary function, whereas 21% were repetitions. The native speakers produced no cases coded as primary repetition, but 36% of their re-accented given items intensified the narrative. The key distinction was whether replacing the repeated element with a pronoun or an unaccented form would reduce meaning. If it would not, the item was coded as repetition; if re-accenting changed the perceived strength, stance, or attainment of a threshold, it was coded as intensification (Eady et al., 1986; Sluijter & van Heuven, 1996; Xu & Xu, 2005). In A's "It BLEW, it BLEW, it BLEW," for example, the

lexical material was repeated, but the second and third accented instances increased the perceived force and duration of the action.³

The learners used this type of prosodically intensifying repetition much less often and therefore had fewer resources for drawing the listener into the unfolding narrative. This contrast captures the difference between audience design and task-completion design: when the listener is not incorporated into the speaker's model, the speaker is more likely to accumulate semantic content in order to "say everything" than to arrange prosodic cues so that the discourse can be followed efficiently (Bell, 1984; Clark, 1996; Lindblom, 1990).

Third, the overall proportion of contrastive functions was similar across groups, but their qualitative realization differed. Learners' contrastive accents were referentially restricted: 6 of the 11 cases referred to the traveller (using traveller or man), 2 to clothes, and only 3 to other referents. Speaker A placed contrastive accents on the verb convinced and the adjective gentle. N's contrastive accent also referred to the depicted participant, described as the windmill keeper, but it was realized through a highly emphatic rhythmic sequence.

(3) But STILL*(H*+L) the WINDMILL(H*+L)// KEEPER(H*+L)//
persevered(H*+L)// enduring(H*+L)// this hardship(H*+L).

In (3), STILL establishes the apex of the contrastive domain. The subsequent word-by-word accented rhythm foregrounds the discrepancy between the windmill keeper's actual persistence and the expected outcome. Another characteristic of learner contrast was the narrowness of the contrastive domain, which often made the contrast appear abrupt and

³ All three instances of blew were realized as L+H*. Their mean F0 values were 293.2 Hz, 273.3 Hz, and 277.6 Hz, and their pitch excursions were 86.4 Hz, 68.2 Hz, and 52.3 Hz, respectively.

insufficiently prepared.

(4) The North Wind realised/ that/ GENTLENESS(H*) is more POWERFUL(H*) than strength(!H*+L).

(5) He took off his jacket and the sun proved(H*+L)// that the GENTLE*(H*) approach is better(H*) than the harsh(H*+L) approach of the wind.⁴

YAQ produced (4) as a single IP. Its emphatic accents were clustered at the end of the IP and were similar in strength, with little evidence of hierarchy. Both prenuclear accents were used contrastively, but no earlier signposting accent prepared the contrastive domain. This differs from the common English pattern in which a prenuclear accent establishes a contrastive dimension and a right-edge nuclear accent supplies resolution (Pierrehumbert & Hirschberg, 1990; Ladd, 2008). The nuclear accent carried new information but was realized as a downstepped falling !H* contour with limited pitch prominence, contributing more to formal prosodic closure than to information-structural closure (Beckman & Ayers Elam, 1997; Ladd, 2008). The contrast was therefore built on two H* accents on given information, while neither the contrastive object nor the outcome was clearly differentiated by prosodic hierarchy. Perception and online-processing research predicts that this combination of multiple undifferentiated strong accents and insufficient signposting increases the burden of referential disambiguation and structural recovery (Birch & Clifton, 1995; Dahan et al., 2002).

Speaker A, by contrast, divided (5) into two IPs. In the first IP, the nuclear accent on

⁴ Every word accompanied by a ToBI label bears emphatic accentuation. To make the contrastive relation visually transparent, only re-accented given items functioning contrastively are capitalized. YAQ's IP had a mean F0 of 256 Hz; the mean F0 values for gentleness and powerful were 301.6 Hz and 287.5 Hz, and the prefallic portion of strength averaged 228 Hz. In A's IP, the mean F0 values for proved, gentle, better, and harsh were 210.2 Hz, 280.6 Hz, 227.8 Hz, and 192.9 Hz, respectively.

proved established a signpost. In the second, the main accent on GENTLE created the contrastive apex; a less prominent H* on better stated the outcome, and the final accent on harsh identified the alternative and closed the unit.

This branching-and-resolution pattern is consistent with a CT–F structure in which a right-edge nuclear accent is supported, where needed, by prenuclear signposting (Büring, 2003, 2016). Because contrastive accents often function as focus, a contrastive domain and a focus domain need not be fully isomorphic. The native speakers established the contrastive domain earlier and then localized the contrastive accent within a focus domain. N’s word-by-word rhythm and A’s preceding IP-level signpost served this purpose. Earlier establishment of contrast, clearer communicative purpose, and a more transparent span of prominence are all characteristic of listener-designed discourse (Gussenhoven, 2004; Ladd, 2008; Watson et al., 2008; Ito & Speer, 2008).

4.3.2 Relations Between Primary and Secondary Functions

The distinction between the two production designs became even clearer when secondary functions were considered. The native speakers frequently used combinations such as signposting → repetition and intensification → contrast. These combinations linked one discourse segment to the next or produced a closing effect, and they worked with the prenuclear–nuclear pattern of anticipation → resolution to maintain a clear hierarchy of signposts. The learners’ strongest combinations were signposting → contrast and contrast → signposting. This suggests that many accents on given information were used to juxtapose visible entities, while coordination between signposting and intensification was comparatively limited. The result was a task-advancing sequence rather than a listener-oriented narrative

design.

When secondary functions were integrated into the primary-function analysis through a controlled weighting procedure, the overall functional distribution remained largely unchanged. The only notable change was that contrast among learners (28.0%) exceeded repetition (26.6%).⁵

The learners' frequent signposting → contrast combination also reveals two distinct problems. Of the 21 tokens in this combination, 20 referred to entities visible in the pictures, including the North Wind, the Sun, the traveller, and the clothes. This distribution suggests that contrast was often built on the repeated marking of visible referents rather than on a clearly organized discourse structure. First, prosodic form did not always match communicative intention. Learners sometimes appeared to intend a contrast, but they failed to realize it at an appropriate position or with an appropriate accentual configuration, reflecting limitations in L2 prosodic proficiency. Second, narrative progression was often achieved by accumulating references to visible objects, which points to a task-completion orientation in discourse design. When the primary objective is to finish the description by saying what is seen, signposting naturally dominates, and contrastive intentions become attached to referential signposts rather than being organized into a listener-guiding discourse structure.

4.4 Summary of the Findings

In addition to the new/given distinction, the analysis examined the distribution of

⁵ Secondary functions were incorporated using a weight of $\alpha = 0.3$ and the formula $\text{Saliency}(f) = [\text{primary-function count} + \alpha \times \text{secondary-function count}] / \text{total number of emphatic accents on given information}$. For native speakers, the resulting percentage-point changes for signposting, contrast, intensification, and repetition were +9.2, +7.9, 0, and +6.6; for learners, they were +1.9, +12.3, 0, and +5.2.

accents across theme and rheme and across transitivity roles—Process, Participant, and Circumstance. The ratio of accented themes to accented rhemes was approximately 1:2 in both groups, with no marked group difference, consistent with the general tendency for the rheme to carry new or focused information (Halliday & Matthiessen, 2014). Clearer differences emerged in transitivity roles. Accents on Processes accounted for an average of 29% of the native-speaker accents but only 17% of the learner accents. Accents on Circumstances accounted for 35% in the native-speaker data and 43% in the learner data. The native speakers thus relied more heavily on accentuation to move events forward, whereas the learners more often accented information about time, place, and manner.

This distribution is compatible with task-oriented semantic accumulation and with UID-based predictions. When the narrative information channel has not been effectively structured, speakers may add more low-efficiency circumstantial material in an attempt to make the description sufficiently explicit (Aylett & Turk, 2004; Jaeger, 2010). Taken together, the findings on IP-level accent configuration, prenuclear–nuclear leverage, and re-accenting of given information indicate that the native speakers’ production was listener-designed, whereas the learners’ production was primarily oriented toward task completion.

5. Discussion

5.1 Sources of the Task-completion Orientation

The data suggest that the Chinese learners approached oral production with a strong task-completion orientation. This pattern should not be attributed to a single cause. Rather, it appears to arise from the combined effects of assessment washback, classroom interactional routines, insufficient instruction in prosody–information-structure relations, cross-linguistic

transfer, and cognitive-load constraints.

First, learners who have long been assessed for grammatical accuracy, lexical appropriateness, and content coverage may internalize a production goal of “saying everything correctly.” In such a learning environment, successful speaking is often understood as the complete and accurate verbalization of required content. This goal may obscure the more fundamental communicative principle that speech is designed for a listener. One consequence is the accumulation of several accented items within a single IP without a clear hierarchy. Instead of using accentuation to guide a listener toward the main point, learners may use it to mark multiple pieces of content that they consider important for task completion.

Second, teacher-fronted interaction in many Chinese EFL classrooms has been shown to be strongly organized around initiation–response–feedback (IRF) sequences (Sinclair & Coulthard, 1975; Li, 2018, 2019). In this interactional routine, learners are frequently positioned as respondents to a teacher’s prompt, and the teacher functions primarily as an evaluator. As a result, learners may come to treat speaking as the delivery of an answer rather than as communication with an addressee (Hall & Walsh, 2002). Their attention is directed toward satisfying task requirements and completing content points, rather than toward designing utterances in ways that make the information structure recoverable for a listener.

Third, the instructional background may also help explain the learners’ weak prosodic organization. Wu (2019, p. 89) observes that “the traditional focus of English phonetic teaching in China has been on the segmental acquisition,” especially the pronunciation of vowels and consonants, whereas suprasegmental knowledge related to speech naturalness, coherence, and understanding has received comparatively limited attention. Learners may therefore receive

training in individual sounds, or in local connected-speech techniques such as linking, reduction, and weak forms, without developing a systematic understanding of how sentence stress and intonation organize information structure. As a result, they may know that certain words can be stressed, but not how accentuation functions to highlight new or contrastive information, mark discourse transitions, or make communicative intention recoverable. This helps explain why their accents often appear dense but weakly organized: prominence is produced, but its discourse function is not always clear (Derwing & Munro, 2005; Levis, 2005; Hahn, 2004; Field, 2005).

Fourth, cross-linguistic transfer may also contribute to the observed pattern. Chinese and English differ substantially in their prosodic organization. Mandarin commonly exhibits post-focus compression, whereas English relies strongly on nuclear accentuation, focus-related pitch accents, and boundary tones to organize information hierarchy (Xu, 1999; Xu & Xu, 2005; Chen, 2008). Without explicit contrastive instruction, Chinese learners may distribute prominence relatively evenly or place it on circumstantial information such as time, place, and manner. In such cases, accentuation is used less to guide the listener toward the intended focus than to cover visible or propositional content.

Finally, cognitive-load constraints should also be considered. Under time pressure, L2 speakers must allocate attention to lexical retrieval, grammatical encoding, pronunciation, fluency, and content organization at the same time. Under time pressure and limited processing resources, learners may allocate attention first to grammatical encoding and propositional completeness. From a limited-attentional-capacity perspective, this may leave less room for prosodic planning and audience design (Skehan, 1998; Skehan & Foster, 2001). This can lead

to speech that is informationally complete but prosodically under-organized.

These institutional, interactional, linguistic, and cognitive pressures jointly produce the task-completion pattern observed in the present study: high accent density, weak focus hierarchy, limited prenuclear–nuclear coordination, and insufficient listener-oriented cues. The outcome is not necessarily unintelligible speech. Rather, it is speech whose communicative organization is less transparent, requiring the listener to do more work in identifying what is new, what is contrastive, and where each unit of information reaches closure.

5.2 Pedagogical Implications

The findings suggest that L2 speaking instruction should move beyond pronunciation accuracy alone and attend more directly to the discourse function of prosody. If learners' difficulty lies not simply in producing stress, but in organizing accentuation in a listener-oriented way, then teaching and assessment should help learners understand how prosody, information structure, and communicative intention work together.

First, speaking assessment should include prosody-information-structure alignment as an explicit criterion. In many assessment rubrics, pronunciation is treated mainly in terms of segmental accuracy, fluency, or general intelligibility. The present findings suggest that assessment should also ask whether the speaker uses accentuation to make the information structure clear. Possible descriptors may include whether the main focus is placed in an appropriate nuclear position, whether IP boundaries support information packaging, whether prenuclear accents prepare rather than compete with nuclear accents, and whether re-accenting of given information serves a recognizable discourse function. In this way, learners are encouraged not only to “speak correctly” but also to “speak so that the listener can follow.”

Second, classroom interaction should provide more opportunities for genuine listener-oriented speaking. If learners are trained mainly through teacher-fronted IRF routines, they may continue to treat oral production as answering a question for evaluation. Classroom tasks should therefore include real or simulated information gaps, peer listeners, role-based narration, and tasks in which the listener must reconstruct information from the speaker's description. For example, in a picture-description task, one student may describe a picture sequence while another arranges the pictures in the correct order. Such tasks make listener understanding a visible outcome of speaking and encourage learners to use prosody to guide attention, mark transitions, and signal narrative closure.

Third, instruction should make the relation between prosody and information structure explicit. Learners need to be shown that English accentuation is not simply a matter of speaking louder or stressing "important words." Rather, it helps organize information into discourse units. Teachers can introduce simple patterns such as given versus new information, contrastive focus, topic-comment structure, and CT-F organization. In particular, CT-F templates can be used to train the sequence of dimension-setting and resolution. Learners can practise how a prenuclear accent establishes a contrastive dimension or discourse frame, while a nuclear accent provides the main resolution or closure. This can help them avoid placing several strong accents in the same IP without functional differentiation.

Fourth, perception-production training should be incorporated. Learners often cannot produce a listener-oriented accentual pattern unless they can first perceive how such a pattern affects interpretation. Teachers may therefore use short paired examples in which the same sentence is produced with different accent placements. Students can compare which version is

easier to understand, which word sounds contrastive, and where the sentence feels complete. They can then imitate, modify, and record their own versions. This type of scaffolded perception–production comparison can reduce online planning demands and improve the match between what learners intend to highlight and what listeners actually perceive.

Fifth, feedback should be made concrete and operational. Instead of giving general comments such as “your stress is unclear” or “your intonation is unnatural,” teachers can refer to specific indicators used in this study. For example, they may ask whether each IP contains a clear main accent, whether the nuclear accent corresponds to the intended focus, whether a prenuclear accent supports the nuclear accent, and whether an accent on given information functions as a signpost, contrast, enhancement, or mere repetition. Student recordings can be used as data for reflective analysis. Learners can mark their own accented words, compare them with peer or teacher annotations, and revise the recording with a clearer listener-oriented design.

Finally, prosodic training should be integrated with discourse-level speaking tasks rather than treated as isolated pronunciation practice. Picture description, storytelling, oral presentation, debate, and problem-solving tasks all provide opportunities to practise accentuation as discourse organization. In these tasks, learners should be encouraged to plan not only what to say, but also how the listener is expected to follow the sequence of information. Pre-speaking planning can include identifying the main point of each IP, deciding which information is new or contrastive, and marking where the narrative should move forward or close. Over time, this can help learners shift from “covering the content” to “organizing understanding.”

Through coordinated changes in assessment, classroom interaction, explicit instruction,

contrastive awareness, perception–production training, and data-based feedback, learners can gradually move from a task-completion orientation toward listener-oriented speech production. Accent and intonation can then assume their full communicative role in organizing information, guiding the listener, and externalizing discourse intention.

6. Conclusion

By comparing emphatic accentuation in picture descriptions produced by native speakers and Chinese learners of English, this study has shown that accentuation, information structure, and discourse intention are systematically related in spoken production. The native speakers generally organized their prosody in a listener-oriented way, using nuclear accents to provide closure within IPs and prenuclear accents, when needed, to establish relevant discourse dimensions. The learners, by contrast, more often produced a linear task-completion pattern, with multiple strong accents occurring within the same IP, weak prosodic closure, and an overuse of signposting. These differences are therefore not merely surface-level pronunciation errors, but different prosodic realizations of communicative planning.

From the perspective of Levelt's (1989) model, the contrast can be understood as arising from different priorities in conceptualization and formulation. The native speakers appeared to plan discourse around a contrastive dimension and a point of resolution, which were then realized through coordinated B-accent and focus-accent patterns. This organization is consistent with Büring's (2003, 2016) account of discourse trees, in which discourse develops through branching relations and reaches closure at relevant nodes. The learners, however, appeared to prioritize propositional completeness. As a result, several semantic points were often mapped onto parallel strong accents without a clear dimension–resolution relation,

leaving listeners to infer both focus structure and prosodic boundaries with fewer guiding cues.

The study contributes to L2 prosody research in three ways. Theoretically, it shows how the distribution of emphatic accents can reveal the link between discourse intention and prosodic realization. Methodologically, it proposes IP–focus alignment, prenuclear–nuclear accent coordination, and the functional distribution of re-accented given information as operational indicators of listener-oriented and task-completion-oriented speech production. Pedagogically, it suggests that spoken English instruction and assessment should attend not only to pronunciation accuracy, but also to how prosody organizes information for a listener.

The conclusions remain provisional because the study was based on a single task and a small sample, especially the two native-speaker comparison participants. Future research should test the proposed indicators across broader discourse contexts, proficiency levels, speaker groups, and larger corpora. Independent annotation and inferential statistical analysis are also needed to further validate the relationship among emphatic accentuation, information structure, and discourse intention.

Funding

This work is supported by the Guangdong Provincial Educational Research Project (Higher Education Science Special Project), “Optimal Analysis of Emphatic-Focus Production Patterns among English Majors” (2022GXJK407).

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional

- explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, 47(1), 31–56.
<https://doi.org/10.1177/00238309040470010201>
- Baumann, S., & Riester, A. (2012). Referential and lexical givenness: Semantic, prosodic, and cognitive aspects. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning* (pp. 119–162). De Gruyter Mouton. <https://doi.org/10.1515/9783110261790.119>
- Beckman, M. E. (1986). *Stress and non-stress accent*. Foris Publications.
<https://doi.org/10.1515/9783110874020>
- Beckman, M. E., & Ayers Elam, G. (1997). *Guidelines for ToBI labelling* (Version 3.0). The Ohio State University.
- Bell, A. (1984). Language style as audience design. *Language in Society*, 13(2), 145–204.
<https://doi.org/10.1017/S004740450001037X>
- Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, 60(1), 92–111. <https://doi.org/10.1016/j.jml.2008.10.001>
- Birch, S. R., & Clifton, C. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech*, 38(4), 365–391.
<https://doi.org/10.1177/002383099503800403>
- Birch, S. R., & Clifton, C. (2002). Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *Journal of Memory and Language*, 47(4), 571–588.
[https://doi.org/10.1016/S0749-596X\(02\)00018-9](https://doi.org/10.1016/S0749-596X(02)00018-9)
- Büring, D. (2003). On D-trees, beans, and B-accent. *Linguistics and Philosophy*, 26(5), 511–545. <https://doi.org/10.1023/A:1025887707652>
- Büring, D. (2016). *Intonation and meaning*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199226677.001.0001>
- Chen, H. (2008). 中国学生英语语调模式研究[On Chinese EFL learners' English intonation patterns]. 上海外语教育出版社[Shanghai Foreign Language Education Press].
- Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. Harper & Row.
- Clark, H. H. (1996). *Using language*. Cambridge University Press.

- Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given–new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. 1–40). Ablex.
- Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, 47(2), 292–314.
[https://doi.org/10.1016/S0749-596X\(02\)00001-3](https://doi.org/10.1016/S0749-596X(02)00001-3)
- Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379–397.
<https://doi.org/10.2307/3588486>
- Eady, S. J., Cooper, W. E., Klouda, G. V., Mueller, P. R., & Lotts, D. W. (1986). Acoustical characteristics of sentential focus: Narrow vs. broad and single vs. dual focus environments. *Language and Speech*, 29(3), 233–251.
<https://doi.org/10.1177/002383098602900304>
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39(3), 399–423. <https://doi.org/10.2307/3588487>
- Gussenhoven, C. (1983). Focus, mode, and the nucleus. *Journal of Linguistics*, 19(2), 377–417.
<https://doi.org/10.1017/S0022226700007799>
- Gussenhoven, C. (1984). *On the grammar and semantics of sentence accents*. Foris.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.
- Gut, U., & Pillai, S. (2014). Prosodic marking of information structure by Malaysian speakers of English. *Studies in Second Language Acquisition*, 36(2), 283–302.
<https://doi.org/10.1017/S0272263113000739>
- Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201–223. <https://doi.org/10.2307/3588378>
- Hall, J. K., & Walsh, M. (2002). Teacher–student interaction and language learning. *Annual Review of Applied Linguistics*, 22, 186–203.
<https://doi.org/10.1017/S0267190502000107>
- Halliday, M. A. K. (1967). Notes on transitivity and theme in English. *Journal of Linguistics*, 3(1), 37–81.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional*

- grammar* (4th ed.). Routledge.
- Heritage, J. (1984). A change-of-state token and aspects of its sequential placement. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 299–345). Cambridge University Press.
- Ito, K., & Speer, S. A. (2008). The role of prosody in the processing of contrastive focus. *Journal of Memory and Language*, 58(4), 689–711.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic processing load. *Cognition*, 116(1), 28–53.
- Jun, S.-A. (2006). Intonational phonology of Korean (revisited). In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 201–229). Oxford University Press.
- Kahnemuyipour, A. (2009). *The syntax of focus*. Oxford University Press.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315.
<https://doi.org/10.1016/j.system.2010.01.005>
- Kjelgaard, M. M., & Speer, S. A. (1999). The role of prosody in sentence processing. *Journal of Memory and Language*, 41(1), 32–58.
- Krivokapić, J. (2007). *Intonational structure and focus* [Unpublished doctoral dissertation]. University of Maryland.
- Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge University Press.
<https://doi.org/10.1017/CBO9780511808814>
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in neural information processing systems 19* (pp. 849–856). MIT Press.
- Li, J. (2018). L1 in the IRF cycle: A case study of Chinese EFL classrooms. *Asian-Pacific Journal of Second and Foreign Language Education*, 3(1).
<https://doi.org/10.1186/s40862-017-0042-y>

- Li, J. (2019). Looking beyond IRF moves in EFL classroom interaction in China. In H. T. Nguyen & T. Malabarba (Eds.), *Conversation analytic perspectives on English language learning, teaching, and testing in global contexts* (pp. 87–109). Multilingual Matters.
- Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Kluwer Academic.
- Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.
- Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). MIT Press.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1(1), 75–116. <https://doi.org/10.1007/BF02342617>
- Schwarzschild, R. (1999). Givenness, AvoidF, and other constraints on the placement of accent. *Natural Language Semantics*, 7(2), 141–177. <https://doi.org/10.1023/A:1008370902407>
- Selkirk, E. O. (1995). Sentence prosody: Intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 550–569). Blackwell.
- Sinclair, J. M., & Coulthard, R. M. (1975). *Towards an analysis of discourse: The English used by teachers and pupils*. Oxford University Press.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Skehan, P., & Foster, P. (2001). Cognition and tasks. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 183–205). Cambridge University Press.
- Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100(4), 2471–2485.
- Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing “given” information: Effects of persistence of grammatical function and surface position. *Language and Speech*, 37(2), 125–145.

- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28(1), 1–30.
- Watson, D. G., Tanenhaus, M. K., & Gunlogson, C. A. (2008). Interpreting pitch accents in online comprehension: H* vs. L+H*. *Cognitive Science*, 32(7), 1232–1244. <https://doi.org/10.1080/03640210802138755>
- Wagner, M., & Watson, D. (2010). The role of prosody in syntactic parsing and sentence comprehension. *Language and Linguistics Compass*, 4(2), 77–89.
- Wu, Y. (2019). Review of Chinese English learners' prosodic acquisition. *English Language Teaching*, 12(8), 89–96. <https://doi.org/10.5539/elt.v12n8p89>
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in speech synthesis. *Journal of the Acoustical Society of America*, 91(3), 1032–1052.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27(1), 55–105. <https://doi.org/10.1006/jpho.1999.0086>
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33(2), 159–197. <https://doi.org/10.1016/j.wocn.2004.11.001>