Prosodic Cues in Relative Clause Attachment Ambiguity Resolution: An Experimental Comparison of English Native Speakers and Chinese EFL Learners

Jiacheng Han

School of Foreign Languages, Qingdao University, Qingdao, China
jerrickhan25@outlook.com

Abstract. This study investigates the relative prosodic cue weighting (intensity, pitch, pause, lengthening) in resolving English relative clause (RC) attachment ambiguity by English native speakers and Chinese EFL learners. A two-alternative forced-choice experiment, combined with a d' analysis to ensure a bias-free measure of perceptual sensitivity, revealed qualitatively different cue-weighting strategies. Native English speakers demonstrated a flexible and sequential strategy, prioritising pitch accent as a decisive feature. In contrast, Chinese EFL learners showed a strong reliance on pause; their accuracy was highest in "pauses alone" conditions, while other prosodic features served as distractors. The two groups also showed opposing biases in standard syntactic analysis (native: low attachment; learners: high attachment). These results provide strong evidence for inappropriate intonation transfer from the native language, with learners inappropriately transferring the functional roles of Chinese tone to English. This study highlights why we need to start teaching the syntactic side of prosody more explicitly in L2 classrooms.

Keywords: prosodic cues, ambiguity resolution, prosodic transfer, relative clause, second language acquisition

1. Introduction

RC attachment ambiguity, for example, in the sentence "Someone shot the servant of the actress who was on the balcony," creates difficulties for sentence processing. Resolving such ambiguities often relies on prosodic cues, the weighting of which varies from language to language [1,2]. Previous studies [3,4] show that when Chinese learners acquire English as a foreign language, they use prosodic cues to mark boundaries and prominence in English, but the relative importance of different cues has not been sufficiently investigated. Given the limited attention paid to prosody in English language teaching in China [5], it is important to understand the mechanism of prosodic transfer. This study investigates how native English speakers and Chinese EFL learners weigh four different prosodic cues to resolve RC ambiguity in English.

2. Literature review

2.1. The Garden Path Model and relative clause attachment

The Garden Path Model [6] is a classic starting point here; it views parsing as a modular, sequential process designed to keep our cognitive load manageable. A central part of this theory is the Late Closure principle [7], which basically argues that we have a universal tendency to attach new information to the most recently processed phrase. In the context of Relative Clauses (RCs), this points toward a "low-attachment" bias. Take the sentence "the daughter of the officer who had the accident" [8] as an example—Late Closure would predict that the officer (NP2) was the one in the accident. However, this "universal" rule doesn't always hold up. A vast body of cross-linguistic research [9] shows that attachment preferences actually vary significantly, which really complicates the idea that RC attachment is governed by syntax alone.

2.2. The syntax-phonology interface

This widespread variation across languages really puts the spotlight on the syntax-phonology interface. It turns out that prosody isn't just an extra; it's a heavy hitter in how we resolve ambiguity [10]. According to frameworks like the Competition Model [11], we're constantly using prosodic cues to steer our parsing decisions. These cues generally fall into two buckets: boundary markers—things like pauses or that slight lengthening before a break—and prominence markers, which are all about pitch and volume [1,12]. We know from previous work [13] that these boundaries shape how we read RCs, while pitch helps us track who or what is being talked about. But the real kicker is that not everyone uses these tools the same way [14]. For instance, Baek [1] noticed that while native English speakers are all about pitch, Korean learners are almost entirely fixated on pauses. ERP data [15] backs this up, showing that Mandarin learners don't process English pitch accents the way natives do, which strongly suggests that our first language really dictates how we weight these prosodic cues.

2.3. Prosodic transfer in L2 relative-clause ambiguity resolution

To make sense of these differences, we have to look at L1 prosodic transfer. It's a classic problem: English is a head-prominence language that uses pitch accents to mark phrases and clear up ambiguity [16]. Mandarin, however, is different—it uses pitch for lexical tones and relies much more on boundary markers for phrasing [4]. The L1 Transfer Hypothesis [17] basically says that we tend to drag our native prosodic habits into any new language we learn, which gets messy when the two languages don't use the same "tools." While we've seen this play out in German-English bilinguals [2], we're still largely in the dark about how Chinese EFL learners specifically handle the prosody of English RCs [3]. This study addresses this gap by asking:

1.What are the default RC attachment preferences for English native speakers and Chinese EFL learners?

2.How do these two groups weight different prosodic cues (pitch, pause, lengthening, intensity) in resolving ambiguity?

3.How do Mandarin prosodic features transfer to English disambiguation strategies?

This study contributes to testing whether theories like the L1 Transfer Hypothesis [17] and L2 Intonation Learning Theory [19] actually hold water in the Mandarin-English context. More than

just a theoretical exercise, this research aims to provide some hard, data-driven evidence that could finally help teachers develop better ways to teach English intonation or prosody.

3. Methodology

This behavioral experiment examined how four prosodic cues influence RC attachment decisions for English native speakers and Mandarin-L1 EFL learners.

3.1. Participants

Fourteen junior English majors from Qingdao, China (6 male; M_age = 21, SD = 1.08) and 12 native American English speakers (7 male; M_age = 24, SD = 2.93) participated.

3.2. Materials

We adapted ten NP1+NP2+RC sentences from the work of Cuetos and Mitchell [8] as our core materials. These were recorded by a male native English speaker in both high- and low-attachment versions. These original, natural recordings—featuring a full range of pitch, intensity, pauses, and lengthening—served as the "All Cues" condition. From this base, we used Praat [20] to generate five more experimental sets. We created a baseline "No Cues" condition by systematically stripping away all four prosodic markers. Then, to test the cues in isolation, we crafted four "single-feature" conditions: Pause-only, Lengthening-only, Pitch-only (using the PSOLA method), and Intensity-only. In these versions, only the target cue was left intact while the others were neutralized. This neutralization was a meticulous process: we flattened pitch contours, cut out silent intervals, shortened nouns to a standard length, and balanced the intensity across both NPs.

3.3. Procedure

We ran the experiment as a two-alternative forced-choice (2AFC) task on the Jinshuju online survey platform [21]. The process was straightforward: participants would listen to a stimulus—such as the sentence about the journalist interviewing "the daughter of the officer who had the accident"—and then immediately answer a question "Who had the accident?". They had to choose between two clear options: "the daughter" or "the officer." To avoid response patterns, we took the 60 target stimuli and pseudo-randomly mixed them with 20 unambiguous filler items. We split these across four experimental lists, so each participant only had to complete one list. Usually, they were able to finish the whole session within ten minutes.

4. Data analysis and results

For the statistical side of things, we used R with the lme4, lmerTest, and emmeans packages [22-24]. Figure 1 gives a clear breakdown of how both groups performed across the various prosodic conditions. What's immediately striking is the divide: while native English speakers are much more accurate with low-attachment readings, the Chinese EFL group shows almost the exact opposite. These learners seem to have a natural lean toward high-attachment interpretations. Furthermore, the data also shows that the learners haven't quite reached the same overall accuracy levels as the native speakers in terms of most conditions.

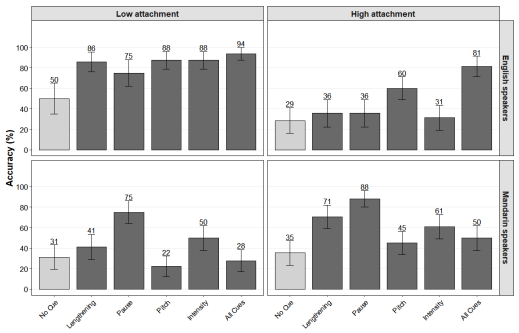

Figure 1. Accuracy rate of responses (%) per prosodic condition

Table 1. Results from the generalized linear mixed-effects model predicting response accuracy (%)

	β	SE	z	p(>|z|)	
(Intercept)	-0.633	0.313	-2.022	0.743	
Language	-0.125	0.313	-0.400	0.689	
Attachment	0.280	0.127	2.212	0.027	*
All Cues condition	1.555	0.458	3.398	0.000	***
Intensity condition	1.014	0.406	2.495	0.012	*
Lengthening condition	1.070	0.414	2.586	0.009	**
Pause condition	1.585	0.442	3.587	0.000	***
Pitch condition	0.902	0.413	2.184	0.029	*
Language:Attachment	-0.680	0.127	-5.349	0.000	***

Language, Attachment, and Prosodic Condition were used as predictors, with the No Cue condition as the baseline. Significance codes: $p < 0.05$, *$p < 0.01$, **$p < 0.001$.

When we ran a GLMM on the trial-level accuracy, two things really stood out: a general advantage for low-attachment and, more importantly, a major interaction between language group and attachment type (see Table 1). This basically confirms the "flip" we saw earlier—native speakers lean toward low attachment, while the learners are pulled toward high attachment. Unsurprisingly, having some prosodic cues was always better than having none at all.

But we didn't just want to look at accuracy; we needed to filter out any simple response bias. So, we turned to d-prime (d') analysis [25]. The LMM results on d' scores (Table 2) highlighted a significant language-by-condition interaction, particularly when it came to Pitch and All Cues. It's clear that native speakers get a much bigger boost from pitch than the learners do.

Table 2. Results from the linear mixed-effects model predicting perceptual sensitivity (d' scores) from Language and Prosodic Condition

	β	SE	t	p(>|z|)	
(Intercept)	-0.363	0.158	-2.306	0.622	
Language	-0.057	0.158	-0.367	0.714	
All cues condition	0.775	0.212	3.655	0.000	***
Intensity condition	0.599	0.212	2.825	0.005	**
Lengthening condition	0.535	0.212	2.525	0.012	*
Pause condition	0.811	0.212	3.830	0.000	***
Pitch condition	0.440	0.212	2.080	0.039	*
Language: All Cues condition	-0.657	0.212	-3.102	0.002	**
Language: Intensity condition	-0.012	0.212	-0.058	0.954	

Table 2. (continued)

Language: Lengthening condition	0.051	0.212	0.242	0.809	
Language: Pause condition	0.335	0.212	1.579	0.117	
Language: Pitch condition	-0.436	0.212	-2.059	0.042	*

Post-hoc pairwise comparisons on d' scores within each group revealed distinct cue-weighting hierarchies. For the native speakers, pitch yields the greatest advantage; their performance with pitch alone was statistically the same as when they had all cues available, and far better than the baseline. Other cues like intensity or pauses helped, but they weren't the main drivers. The Chinese learners, however, had a completely different pattern. For them, the pause was the only cue that truly mattered. In fact—and this is interesting—they performed better with pauses alone than with the full cue set. It seems that adding pitch, intensity, or lengthening actually acted as a distraction, or "noise," for them. In those other conditions, their accuracy didn't even budge from the baseline, suggesting they were relying almost exclusively on the pause.

5. Discussion

5.1. Attachment preference in each language group

Our findings point to a clear split in how these two groups default to different parsing strategies. For the native English speakers, the preference for low attachment was exactly what we'd expect—it mirrors the Late Closure principle [7] and fits right in with what's been reported in the literature [1,26]. On the other hand, the Chinese EFL learners went the other way. Their tendency toward high attachment echoes previous studies like Dai [27]. This isn't just a random choice; it likely speaks to the influence of their L1 syntactic background or perhaps the unique way they've navigated their L2 learning journey [26].

5.2. Group differences in prosodic cue weighting

The real "story" of the data, however, lies in how these groups weigh different prosodic cues. Native English speakers seem to have a very adaptable hierarchy. For them, pitch is the heavy lifter; it's so effective that they can almost rely on it alone to understand the syntax, which makes sense given that English is a head-prominence language [16].

The Chinese learners, by contrast, seem stuck on pauses. Their reliance on this single cue is so rigid that adding more information—like pitch or intensity—actually made their performance worse. This is a fascinating bit of interference. If we look at this through the lens of the Cognitive Load Hypothesis [28], it suggests that trying to process multiple, unfamiliar L2 prosodic dimensions simply overloads their working memory. A pause, being a very salient and "one-dimensional" marker, is much easier for them to grab onto, especially since it aligns with how Mandarin handles phrasing [29]. It's a perfect example of what LILt [19] predicts: when L2 features drift too far from the L1 "map," they become incredibly hard to master.

5.3. The effect of L1 prosodic transfer

All of this provides a strong case for what I'd call "maladaptive" L1 transfer. In Mandarin, pitch is busy doing lexical work (tone), so it's not really available to signal phrasal boundaries; that job falls to pauses [30]. It seems learners are dragging this functional mapping directly into English. They treat the most crucial English intonation cues as mere background noise while over-relying on

pauses to do all the structural heavy lifting. This "misapplication" of L1 habits explains not just why their strategy is so different, but also why they struggle so much more than native speakers when the prosodic signals get complex.

6. Conclusion and implications

Ultimately, this study underscores a fundamental difference in processing: native speakers lean on pitch, while Chinese learners are stuck on pauses. This reliance on pauses is a direct carryover from Mandarin, and it often acts as a barrier to picking up the more subtle pitch cues in English. These findings send a clear message to educators: it's time to look beyond vocabulary and individual sounds. If we want learners to truly master English, we need to explicitly teach the structural functions of intonation. With the right kind of perceptual training, we can help them overcome these native-language biases and move toward a more native-like understanding of English syntax.

References

[1] Baek, H. (2022). Prosodic cue weighting in the processing of ambiguous sentences by English native listeners and Korean listeners of English. The Journal of the Acoustical Society of America, 151(1), 158–167.

[2] O'Brien, M. G., Jackson, C. N., & Gardner, C. E. (2014). Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English. Applied Psycholinguistics, 35(1), 27–70.

[3] Xue, L. Y. (2015) 'Yunlü xiansuo zai dier yu kouyu jiagong zhong de shiyong' [Use of prosodic cues in second-language spoken processing] (Doctoral thesis, Zhejiang University).

[4] Xue, L. & Yue, M. (2024). Discrepancy in prosodic disambiguation strategies between Chinese EFL learners and English native speakers. International Review of Applied Linguistics in Language Teaching, 62(2), 1065–1092. https: //doi.org/10.1515/iral-2022-0004

[5] Zhu, Y. H. (2024). 'Zhongguo Yingyu Xuexizhe Yingyu Yinhang Zhongzhong Ganzhi he Chanchu Qingkuang ji Xiangying Yunlü Jiaoxue Celüe Yanjiu' [Perception and production of English pitch accent by Chinese learners and corresponding prosodic teaching strategies], Yingyu Guangchang [English Square], (27), 90–93. doi: 10.16723/j.cnki.yygc.2024.27.027

[6] Frazier, L. (1979). On comprehending sentences: Syntactic parsing strategies. University of Connecticut.

[7] Frazier, L. (1987). Theories of sentence processing. Cambridge University Press.

[8] Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. Cognition, 30(1), 73–105.

[9] Jun, S. A. (2003). Prosodic phrasing and attachment preferences. Journal of Psycholinguistic Research, 32, 219–249.

[10] DeDe, G. (2010). Utilization of prosodic information in syntactic ambiguity resolution. Journal of Psycholinguistic Research, 39, 345–374.

[11] MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? Journal of Memory and Language, 28(3), 255–271.

[12] Nespor, M., & Vogel, I. (2007). Prosodic phonology: With a new foreword (Vol. 28). Walter de Gruyter.

[13] Arnold, J. E. (2008). THE BACON not the bacon: How children and adults understand accented and unaccented noun phrases. Cognition, 108(1), 69–77.

[14] Mergen, F., & Karakoç, N. Y. (2024). Relative clause attachment preferences of late bilinguals. International Journal of Bilingualism. https: //doi.org/10.1177/13670069241270762

[15] Lee, A., Perdomo, M., & Kaan, E. (2020). Native and second-language processing of contrastive pitch accent: An ERP study. Second Language Research, 36(4), 503–528.

[16] Jun, S. A. (Ed.). (2006). Prosodic typology: The phonology of intonation and phrasing (Vol. 1). Oxford University Press.

[17] Flege, J. E., Munro, M. J., & MacKay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. The Journal of the Acoustical Society of America, 97(5), 3125–3134.

[18] Morrill, T. H. (2016). A facilitating effect of a prosodic transfer on non-native fluent speech listening. Language, Cognition and Neuroscience, 31(6), 801–816.

[19] Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations (pp. 171–188). Springer Berlin Heidelberg.
[20] Boersma, P., & Weenink, D. (2005). Praat. Doing phonetics by computer (Version 5.1).
[21] Beijing Jinshuju Technology Co., Ltd. (2024). Jinshuju [Online survey platform]. https: //jinshuju.net/
[22] Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., … & Bolker, M. B. (2015). Package 'lme4'. convergence, 12(1), 2.
[23] Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. Journal of Statistical Software, 82, 1–26.
[24] Lenth, R., & Lenth, M. R. (2018). Package 'lsmeans'. The American Statistician, 34(4), 216–221.
[25] Macmillan, N. A., & Creelman, C. D. (2005). Detection theory: A user's guide.
[26] Dai, Y. (2016). Ambiguity resolution of English relative-clause attachment by Chinese learners. Journal of Foreign Languages, 39(6), 65–74.
[27] Yao, B., & Scheepers, C. (2018). Direct speech quotations promote low relative-clause attachment in silent reading of English. Cognition, 176, 248–254.
[28] Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285.
[29] Kuang, J., Chan, M. P. Y., Rhee, N., Liberman, M., & Ding, H. (2022). The mapping between syntactic and prosodic phrasing in English and Mandarin. Interspeech, 3443–3447.
[30] Li, W., & Yang, Y. (2009). Perception of prosodic hierarchical boundaries in Mandarin Chinese sentences. Neuroscience, 158(4), 1416–1425.