Asymmetric Efficiency of Juncture Perception in L1 and L2

Martin Ho Kwan Ip, Anne Cutler

The MARCS Institute and ARC Centre of Excellence for the Dynamics of Language,
Western Sydney University, Penrith South, NSW 2751, Australia
`m.ip@westernsydney.edu.au, a.cutler@westernsydney.edu.au`

Abstract

In two experiments, Mandarin listeners resolved potential syntactic ambiguities in spoken utterances in (a) their native language (L1) and (b) English which they had learned as a second language (L2). A new disambiguation task was used, requiring speeded responses to select the correct meaning for structurally ambiguous sentences. Importantly, the ambiguities used in the study are identical in Mandarin and in English, and production data show that prosodic disambiguation of this type of ambiguity is also realised very similarly in the two languages. The perceptual results here showed however that listeners' response patterns differed for L1 and L2, although there was a significant increase in similarity between the two response patterns with increasing exposure to the L2. Thus identical ambiguity and comparable disambiguation patterns in L1 and L2 do not lead to immediate application of the appropriate L1 listening strategy to L2; instead, it appears that such a strategy may have to be learned anew for the L2.

Index Terms: prosody, prosodic juncture, boundary, syntactic ambiguity, speech perception

1. Introduction

Human listeners use prosody to segment speech streams into meaningful units, as attested by evidence from infancy to old age and across many languages [1-4]. Indeed, it has been proposed that the role of prosody in signaling juncture is universal [5]. A universal account is supported by evidence that listeners can use prosody to judge the presence of a boundary in an unfamiliar non-native language [6,7].

However, even if there is a common universal substrate that dictates the way we process prosodic junctures (thus, in both a first [L1] and second language [L2]), this universal substrate may, over the course of development, be gradually shaped by the structure of our mother tongue, leading to prosodic processing strategies that are particularly optimised for L1 [8]. On such an account, it is still an empirical question whether listeners can generalise prosodic strategies to process relevant boundaries in L2; L1 strategies may still be used even with some fluency in L2, because acquisition of L2 prosody is a protracted process [9] and learners rarely attain a native-like level of prosodic production [10].

On the other hand, L2 production data reveal that speakers do not fully transfer their L1 cues to syntactic structure, but can exhibit appropriate L2 cues quite early in learning. For example, prosodic disambiguation in German by English L2 learners of German, and in English by German L2 learners of English, proved in each case to resemble the L2 target cues rather than the cues in the speakers' own L1 [11]. Similarly, fourth-semester L2 French learners correctly produce L2 prosodic cues to relative-clause attachment ambiguity [12].

Although the literature contains extensive data on L1 and L2 juncture perception, involving many languages, the L1 and L2 prosodic cues in question can differ extensively. Thus studies comparing L1 and L2 prosodic disambiguation often involved languages with different prosodic realisation of boundary cues (e.g., in [11], English disambiguation involved only pitch accent, while the German disambiguation involved both pitch accent and F_0 rise).

The L1 versus L2 experiments we report here, in contrast, compare Mandarin and English. In these two languages, despite their typological distance, ambiguous utterances allow exactly the same kind of syntactic ambiguity. Consider the following examples:

(a) 姥姥 / 给 / 她 # 狗　肉 / 吃
   Grandma / gave / her #　dog meat / to eat

(b) 姥姥 / 　给 / 她 / 狗 # 肉 / 吃
   Grandma / gave / her / dog # meat / to eat

The two sentences differ only in the juncture location. In (a), the juncture (#) is realised earlier on in the utterance, giving a sentence with a feminine personal pronoun (her/她) as the indirect object and a compound noun as the direct object (dog meat/狗肉). In (b), the same (and segmentally identical) sentence is produced with a later boundary, after dog/狗, so that the personal pronoun becomes a possessive determiner.

In an unpublished crosslanguage production study conducted in our laboratory, we identified two main ways in which speakers in both languages can use prosody to realise syntactic boundaries. First, speakers in both languages would insert a pause at the boundary, and second, they would lengthen the vowels immediately preceding the boundary. Prosodic disambiguation can also be achieved through language-specific changes in F_0 contours (e.g., boundary tones [13], tone sandhi [14]), but this cue tends to vary both across and within speakers [15], and studies in both languages suggest that boundaries are most reliably signalled by the duration adjustments that we also observed [16, 17, 18].

Given the identical ambiguous structures and the same preferences for juncture production, it would be reasonable to expect that speakers of one of these languages would transfer their L1 perception strategies to the other language as L2. We tested this with Mandarin native speakers listening to their L1 and to L2 English. Our participants listened to syntactically ambiguous sentences in their L1 and L2, and in a novel disambiguation task had to determine the intended meanings as quickly as possible. If the Mandarin speakers can draw on their L1 experience in processing the L2, then they should show similar response patterns and accuracy rates in both languages. If L1-optimised prosodic processing requires the presence of L1 speech, however, a different result may ensue.

2. L1 Experiment

2.1. Method

2.1.1. Participants

The sample comprised 40 native speakers of Mandarin (M_{age} = 25.12 years, SD = 3.61 years, range: 18.75-38.30 years; 21 females). All participants were born in Mainland China and had been living in Australia for an average of 1.86 years (SD = 2.27 years, range: 8 days-10 years).

2.1.2. Materials

Twenty-two syntactically ambiguous experimental sentences were chosen, each having two different interpretations resulting from different juncture placement (see Figure 1). The sentences were recorded in their two versions by a female native speaker at a natural fast-normal rate. In the first version, the speaker produced a sentence with an "Early Juncture", where the boundary occurred earlier in the utterance (e.g., 刘波不小心给她#老鼠药吃 "*Liubo accidentally gave her # rat poison*"). In the second version, the same segmentally identical sentence was produced with a "Late Juncture", where the boundary occurred later in the utterance (e.g., 刘波不小心给她老鼠#药吃 "*Liubo accidentally gave her rat # poison*"). For each experimental sentence, the speaker also produced a pair of interpretation sentences that corresponded to the intended meaning of the "Early" and "Late Juncture" versions (e.g., translated in English, "*Liubo gave rat poison to Zhenni*" vs. "*Liubo gave rat poison to Zhenni's pet rat*).

Twelve additional filler sentences and their corresponding pair of interpretation sentences were also recorded. There were thus two counterbalanced experimental conditions, each containing one juncture version of each of the 22 experimental sentences, plus the additional 12 filler sentences. The speaker was asked to produce each version of the experimental sentences in a way that would match its corresponding interpretation sentence.

2.1.3. Procedures

The disambiguation task was administered using E-Prime software on a laptop computer, with attached to it a set of headphones and a Chronos USB-based device for button pressing. All instructions were given in the form of a pre-recorded voiceover script made by the same speaker who produced the stimuli. At the start of each trial, participants saw on their screen two interpretation sentences that corresponded to the left and right buttons in front of them. Participants heard the test sentences and were required to choose for each sentence its intended meaning, by pressing the button that matched the correct interpretation sentence. All participants were asked to "pay careful attention to the meaning of each sentence" and choose the correct button "as soon as they understood the sentence". We recorded participants' response times and number of correct responses. All participants correctly disambiguated at least 64% of the experimental sentences. None of the participants received any explicit instructions on how to disambiguate the sentences.

At the end, all participants completed a recognition test to judge whether each of the 22 sentences in the list were from the experiment. Half of these sentences were from the experiment. All participants scored above 16 out of 22 (73%) on the test (M = 90.68%, SD = 8.17%, range: 73-100%).

2.2. Results and Discussion

More than 90% of participants' correct responses were made by pressing the button after the test sentence was played. Therefore, we measured response times (RT) as the latency duration between the offset of the test sentence and participants' button presses. Only data for accurate disambiguations were included in our analyses.

Participants showed significantly faster disambiguation RT to "Early Juncture" sentences (M = 1072.15 ms, SD = 427.44 ms) compared to "Late Juncture" sentences (M 1219.05 ms, SD = 549.68 ms), $t(1, 39)$ = 2.67, p = .011, d = .422. For accuracy, participants had an average of 3.28 incorrect responses out of 22 sentences. Both the "Early" and "Late Juncture" sentences had a similar number of errors.

We also conducted acoustic analyses of the stimuli to determine the prosodic features that might have contributed to the RT differences in the two juncture versions. For each sentence, we measured the pause and pre-boundary vowel durations of the boundary that would indicate an early juncture and the same duration cues that would indicate a late juncture.

liu2	po1	pu4	ɕiao3	ɕin1	kei3	tʰa1	lau3	ʂu3	jau4	tɕʰi1
刘	波	不	小	心	给	她	老鼠		药	吃
Liubo		accidentally			gave	her	rat		poison	to eat

(a) 刘波不小心给她 # 老鼠药吃
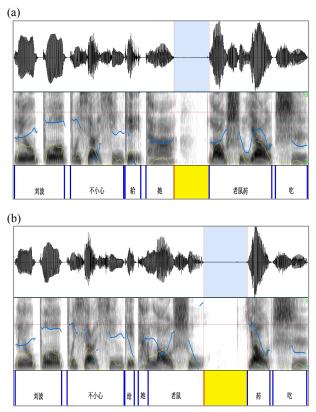
(b) 刘波不小心给她老鼠 # 药吃

(a)

(b)

Figure 1: *Waveforms and spectrograms of an example experimental sentence in the "Early Juncture" (a) and "Late Juncture" (b) version. The blue/yellow shaded portion represents the pause duration of the juncture.*

As revealed in Figure 1, all sentences displayed significantly longer pauses and pre-boundary vowels at their designated boundaries. However, "Late Juncture" sentences were found to have greater increases in pause duration compared to "Early Juncture" versions. This suggests that there was still some variation in the extent to which duration cues were enhanced in the different juncture versions, even though the overall pattern of acoustic features was the same.

In the light of the acoustic data, the slower RT in the "Late Juncture" sentences may indicate that listeners were paying attention to the extra increases in the pause duration. At the same time however, the slower RT might have also been due to the late arrival of the boundary pause. In the second experiment, we tested whether native Mandarin speakers would adopt the same prosodic strategies to process the same kind of structural ambiguity in English. If listeners show the same processing strategies in a second language, then they should show comparable accuracy scores and a similar pattern of RT in the "Early" and "Late Juncture" sentences.

3. L2 Experiment

3.1. Method

3.1.1. Participants

There were 36 native speakers of Mandarin Chinese (M_{age} = 26.36 years, SD = 4.92 years, range: 20.73-43.62 years; 19 females) who had been living in Australia for any period between 28 days to 24.77 years (M = 2.80 years, SD = 4.73 years). Due to recruitment constraints, most of these participants were those who had already participated in the previous experiment. We excluded data from 7 participants who failed to disambiguate at least 64% of the experimental sentences, and 3 participants who scored below 16 out of 22 on the recognition test, leaving a remaining total of 26 participants in the final sample.

In the control group, there were a total of 40 native English speakers (M_{age} = 22.50 years, SD = 7.70 years, range: 17.89 – 53.50 years; 31 females).

3.1.2. Materials and Procedures

The procedures were identical to those in the previous experiments, only this time, the sentences were in English. The English stimuli were recorded by a native female speaker of Australian English and were comparable with the Mandarin sentences in terms of their structural ambiguity. On average, the Mandarin-speaking participants scored 19.61 out of 22 (89.14%) in the recognition test (SD = 6.69%, range: 73-100%), which was not significantly different from that of the English speakers (M = 88.64%, SD = 9.14%, range: 73-100%) and their recognition scores in the L1 Experiment.

3.2. Results and Discussion

Analyses show that Mandarin listeners did fully not transfer their L1 prosodic strategies in a second language. First, there was no significant difference in RT between the "Early" (M = 1332.22 ms, SD = 514.67 ms) and "Late Juncture" sentences (M = 1309.35 ms, SD = 672.87 ms), $t(1, 28) = 0.23$, p = .822. On the other hand, the English native speakers showed significantly faster RT in "Late Juncture" sentences (M = 1109.43 ms, SD = 555.64 ms) compared to the "Early Juncture" sentences (M = 1355.63 ms, SD = 704.88

ms), $t(1, 39) = 3.59$, p = .001, d = .568. Similarly, the Mandarin speakers had, on average, 5.62 incorrect responses out of 22 English sentences, significantly more than the errors found in the L1 experiment, $t = 30.83$, $p < .001$.

Similar to the Mandarin stimuli, the English sentences found longer pauses and pre-boundary vowels in both versions. However, the English stimuli did not show a greater increase in pause duration in the "Late Juncture" versions, and unlike Mandarin, the "Late Juncture" English sentences showed an extra increase in pre-boundary vowel duration.
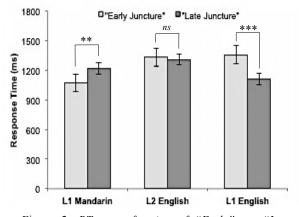

Figure 2: *RT as a function of "Early" vs. "Late Juncture" versions in L1 Mandarin (L1 Experiment) and L2 English by native Mandarin speakers and L1 English by native English speakers (L2 Experiment).*

3.2.1. Length of Stay and L2 Disambiguation

As our participants were not fully uniform with respect to how long they had spent in non-Mandarin-speaking environments, an additional analysis was conducted to assess whether participants' RT was related to their exposure to English as a foreign language while living in Australia. Participants' difference scores in RT were calculated by subtracting their average RT in "Early Juncture" sentences from the RT in "Late Juncture" sentences. A Pearson's correlational analysis was performed to calculate the association between participants' RT difference score and their length of stay in an English-speaking country, and the result showed a significant positive correlation, r = .42, p = .035 (see Figure 3 below).
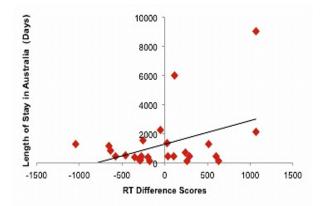

Figure 3: *Significant positive correlation between participants' length of stay (i.e., date of testing minus date of arrival) in an English-speaking country (in days) and their RT difference scores in English.*

4. General Discussion

The present experiment offers a useful insight into how listeners use prosody to disambiguate native and nonnative speech. Using L1 Mandarin and L2 English sentences that involve the same structural ambiguity and the same means of prosodic production, our RT and accuracy findings show that Mandarin listeners do not fully transfer their L1 strategies to disambiguate the sentences in English, although they could have done so and achieved an efficient perceptual outcome. However, it is noteworthy that their response pattern to "Early" and "Late Juncture" sentences in English was also not comparable with that of the native English speakers.

One possible explanation for the lack of L1 to L2 transfer could be that the English and Mandarin stimuli exhibited different degrees of duration increases. "Late Juncture" sentences in Mandarin showed a greater increase in pause duration, while in English there was instead an extra increase in pre-boundary vowel duration. This may partly suggest that Mandarin speakers have a delayed RT in Mandarin "Late Juncture" sentences because they were paying more attention to the extra increases in pause duration, while the faster disambiguation RT in the same version in English could be facilitated by the duration increase in pre-boundary vowels. This suggests that a language-specific preference for a given prosodic cue to boundary placement (e.g., durational cues; pitch accents) is far from the whole story; the precise details of a cue's realisation are also part of the L1 strategy.

Indeed, there is considerable evidence [19, 20, 21] that even when the same cues (e.g., VOT, domain-initial strengthening) are used across languages, the exact realisation may vary. However, it is important to note that the lack of L1 to L2 transfer in our experiments cannot be fully explained by duration adjustment differences in the English vs. Mandarin stimuli. There was a significant positive association between the Mandarin speakers' length of stay in Australia and the degree to which their disambiguation RT in L2 reflected the same RT pattern found in L1 (i.e., a slower RT in "Late Juncture" compared to "Early Juncture" sentences). This unexpected finding indicates that longer time spent in a non-native environment increases the chance of L1 to L2 transfer of disambiguation strategies. Future research should further investigate whether this link might have been mediated by other factors (e.g., age of arrival, rather than length of stay). In any case, our results indicate that native Mandarin listeners' differing response patterns in L1 and L2 sentences, despite similarities in syntactic ambiguity and prosodic production, reflect more than just the small differences in duration increases in the stimuli.

Why did the Mandarin listeners fail to exhibit the same response patterns and accuracy rate in L2? One reason could be that disambiguation strategies are indeed specifically tailored to L1 processing. Then it could be the case that it takes time to learn how to assess relative duration as realised on a new (L2) segmental repertoire. Alternatively, listeners may gradually learn to concentrate on those prosodic dimensions that are more reliably related to the boundary occurrence in their native language [22]. Finally, however, the lack of L1 to L2 transfer revealed in our experiments may demonstrate that L1 prosodic disambiguation is actually learned as a purely language-specific strategy, and as a result all learners must learn from scratch the prosodic system of their L2. Further studies will be needed to decide between these alternatives.

From a methodological standpoint, the present experiments provide a novel and quite parsimonious approach to examining listeners' prosodic disambiguation of ambiguous utterances. A final question that warrants further research is the degree to which listeners' use of these prosodic cues is on-line. In recent years, increasing research has prioritised global patterns of prosodic phrasing, as opposed to the size of local cues [23]. To address this question, future research could examine whether listeners show a drastic decrease in accurate disambiguation after the duration of the critical juncture is rendered uninformative.

5. Conclusions

Our findings provide evidence that listeners do not generalise their L1 prosodic disambiguation strategies to L2, even though both the L1 and L2 speech involved the same syntactic ambiguity and similar prosodic means of disambiguation. The transfer of an effective L1 cue to an L2 in which it would be equally effective still requires an explicit learning process.

6. Acknowledgements

7. References

[1] M. Soderstrom, A. Seidl, D. G. Kemler Nelson, & P. Jusczyk, "The prosodic bootstrapping of phrases: Evidence from prelinguistic infants," *Journal of Memory and Language*, vol. 49, no. 2, pp. 249-267, 2003.

[2] A. Gout, A. Christophe, & J. L. Morgan, "Phonological phrase boundaries constrain lexical access II. Infant data," *Journal of Memory and Language*, vol. 51, no. 4, pp. 548-567, 2004.

[3] K. J. Hoyte, H. Brownell, & A. Wingfield, "Components of speech prosody and their use in detection of syntactic structure by older adults," *Experimental Aging Research*, vol. 35, no. 1, pp. 129-151, 2009.

[4] P. Keating, T. Cho, C. Fougeron, & C-S. Hsu, "Domain-initial articulatory strengthening in four languages," in *Laboratory Phonology VI: Phonetic Interpretation*, J. Local, R. Ogden, & R. Temple (eds.). Cambridge: Cambridge University Press, pp. 145-163, 2003.

[5] D. L. Bolinger, "Intonation across languages," in *Universals of human language II: Phonology*, J. Greenberg (ed.). Palo Alto, CA: Stanford University Press, pp. 471-524, 1978.

[6] R. Carlson, J. Hirschberg, & M. Swert, "Cues to upcoming Swedish boundaries: Subjective judgment studies and acoustic correlates," *Speech Communication,* vol. 46, pp. 326-333, 2005.

[7] A. D. Endress, & M. D. Hauser, "Word segmentation with universal prosodic cues," *Cognitive Psychology*, vol. 61, no. 2, pp. 177-199, 2010.

[8] A. Cutler, *Native Listening*. Cambridge, MA: MIT Press, 2012.

[9] M. C. Pennington, & N. E. Ellis, "Cantonese speakers' memory for English sentences with prosodic cues," *Modern Language Journal*, vol. 84, pp. 372-389, 2000.

[10] I. Mennon, "Bi-directional interference in the intonation of Dutch speakers of Greek," *Journal of Phonetics*, vol. 32, no. 4, pp. 543-563, 2004.

[11] M. G. O'Brien, C. N. Jackson, & C. E. Gardner, "Cross-linguistic differences in prosodic cues to syntactic disambiguation in German and English," *Applied Psycholinguistics*, vol. 35, no. 1, pp. 27-70, 2014.

[12] L. Dekydtspotter, B. Donaldson, A. C. Edmonds, A. L. Fultz, & R. A. Petrusch, "Syntactic and prosodic computations in the resolution of relative clause attachment ambiguity by English-French learners," *Studies in Second Language Acquisition*, vol. 30, no. 4, pp. 453-480, 2008.

[13] D. Watson, & E. Gibson, "Intonational phrasing and constituency in language production and comprehension," *Studia Linguistica*, vol. 59, pp. 279-300, 2005.

[14] S. R. Speer, C-L. Shih, & M. L. Slowiaczek, "Prosodic structure in language understanding: Evidence from tone sandhi," *Language and Speech*, vol. 32, no. 4, pp. 337-354, 1989.

[15] A. J. Schafer, S. R. Speer, P. Warren, & S. D. White, "Intonational disambiguation in sentence production and comprehension," *Journal of Psycholinguistic Research*, vol. 29, no. 2, pp. 169-182, 2000.

[16] D. H. Klatt, "Vowel lengthening is syntactically determined in a connected discourse," *Journal of Phonetics*, vol. 3, pp. 129-140, 1975

[17] I. Lehiste, J. P. Olive, & L. A. Streeter, "Role of duration in disambiguating syntactically ambiguous sentences," *Journal of the Acoustical Society of America*, vol. 60, no. 5, pp. 1199-1202, 1976.

[18] B. Wang, Y. Xu, & Q. Ding, "Interactive prosodic marking of focus, boundary and newness in Mandarin," *Phonetica*, vol. 75, no. 1, pp. 24-56, 2018.

[19] J. B., Pierrehumbert, & D. Talkin, "Lenition of /h/ and glottal stop," in *Papers in Laboratory Phonology II*, G. J. Docherty, & D. R. Ladd (eds.). Cambridge, Cambridge University Press, pp. 90-117, 1992.

[20] T. Cho, & J. M. McQueen, "Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress," *Journal of Phonetics*, vol. 33, no. 2, pp. 121-157, 2005.

[21] C. Kuzla, & M. Ernestus, "Prosodic conditioning of phonetic detail in German plosives," *Journal of Phonetics*, vol. 39, no. 2, pp. 143-155, 2011.

[22] K. B. Shatzman, & J. M. McQueen, "Segment duration as a cue to word boundaries in spoken-word recognition," *Perception and Psychophysics,* vol. 68, no. 1, pp. 1-16, 2006.

[23] L. Frazier, K. Carlson, & C. Clifton Jr, "Prosodic phrasing is central to language comprehension," *Trends in Cognitive Sciences*, vol. 10, no. 6, pp. 244-249, 2006.