Effects of part of speech: Primitive or derived from word frequency?

Chengxia Wang, Yi Xu
Dept. of Speech, Hearing and Phonetic Sciences, University College London, UK

Abstract

Part of speech (POS hereafter) is known to affect both duration and F0, such that, nouns are longer and higher in F0 than verbs. In this study we tested the hypothesis that the POS effects are actually a word frequency effect, and that this effect is predictable from information theory. We tested this hypothesis by comparing 44 phonologically matched noun-verb pairs in Mandarin. Results show that there were clear effects of word frequency on duration, but no effects on F0. In contrast, no effects of POS were found on either duration or F0. We conclude that there are no primitive POS effects on duration or F0, but the frequency effect on duration may lead to a weak POS effect given sufficient corpus size.

Key words: part of speech effect, frequency effect, information theory, duration, F0

Introduction

In both recognition and synthesis modeling, part of speech (POS) is taken as an essential input feature during training. But it is unclear why it is so important. Some informal research reports that nouns are longer than verbs (Lightfoot 1970, Coker et al. 1973). However, those studies did not take phrasing and word frequency into account. Word frequency is already known to have an effect on some acoustic parameters（Xu et al. 1973, Aylett et al. 2004, 2006）. So it is possible that POS effects are a result of word frequency. Frequency of occurrence is also important in information theory, as it is directly related to information load (Shannon 2001).

In the present study, we seek to establish a clear link between POS effects and word frequency. Our hypothesis is that the POS effect on duration is derivable from a word frequency effect, such that the latter is much more robust than the former. This is done by comparing pairs of nouns and verbs in Mandarin that are phonetically identical.

Method

Stimuli

44 pairs of sentences were composed, each containing a pair of noun-verb homophones, 12 of which were monosyllabic words and 32 were disyllabic words. Each pair of sentences shared the same total length and position and phrasing of the target words. This design therefore tightly controlled the effects of syllable structure, phrasing and sentence length. The word identity

(hence POS) was represented by their corresponding Chinese characters. The nouns were all made-up first names. There was no word stress or focus in the target words. Frequencies of the verbs were taken from Modern Chinese Frequency Dictionary (Beijing Language and Culture University 1986), while the frequencies of the nouns (names) were calculated from an online name dictionary (online 2016).

Participant

Five female and four male native Mandarin speakers were recorded. They were all college students studying in London and were born and raised in Beijing. The recordings were conducted in a quiet room in Scape Shoreditch, London.

Recording Procedures

All stimulus sentences were presented in Chinese on a computer screen in a random order, with three repetitions each (in separate blocks). In total, 88 (sentences) × 9 (subjects) × 3 (repetitions) = 2376 sentences were recorded.

Categorization of word frequency

To be able to compare word frequency effect directly with POS effect, we transformed the gradient frequency into a categorical variable, using the median frequency as the dividing line. This way, each target word was treated as either high frequency or low frequency.

Results

For the POS effect, verbs are shorter than nouns in terms of mean duration. Two one-way ANOVAs show that POS has a significant effect on duration, $F(1,22) = 5.069$, $p = .035$ for monosyllabic words, and $F(1,62) = 12.691$, $p = 0.001$ for disyllabic words.

For the word frequency effect, low frequency words are longer than high frequency words. However, one-way ANOVAs shows no significant effect on monosyllabic words ($F(1,22) = 1.446$, $p = .242$), but a significant effect on disyllabic words ($F(1,62) = 17.477$, $p < 0.001$).

Because POS and frequency affect duration in the same direction, they are potentially confounded. To avoid confounding, two ANCOVAs were conducted with POS as independent variable, frequency as covariate and duration as dependent variable. The results are displayed in Table 1.

Table 1. ANCOVA results for duration (ms) of monosyllabic and disyllabic words.

Levels	Monosyllabic		Disyllabic	
	POS	frequency	POS	frequency
verb/high frequency	170.7	179	328.7	323
noun/low frequency	198.7	196	367.5	367.6
F value	3.559	0.272	3.429	7.532
p value	0.073	0.608	0.069	0.008

For monosyllabic words, neither POS nor frequency has a significant effect on word duration. A possible cause of the disappearance of POS effect is the small amount of data. But an additional factor is that the division of the words into high and low frequency ones mixed up the syllable structures, so that the frequency comparison was no longer based on minimal pairs of homophones.

For disyllabic words, the POS effect is no longer significant, but the frequency effect is highly significant. This may have been due to several causes: (1) there were more disyllabic than monosyllabic words, (2) some high frequency words had no homophone counterparts with low frequency, and (3) frequency did affect duration significantly. To rule out the second possibility, we looked into each homophone pair of high and low frequency words. If a low frequency noun-verb pair had a complex syllable structure, which may increase duration, it was deleted. After the deletion, 27 pairs were left. Again the frequency effect was significant while the POS effect was not (Table 2).

Table 2. ANCOVA results for duration (ms) of disyllabic words.

Levels	POS	frequency
verb/high frequency	322	319
noun/low frequency	363	364
F value	2.286	5.241
p value	0.137	0.026

Also tones can affect F_0, so sentence pairs in which target words did not have identical tones were then excluded from the data, leaving 12 pairs of sentences with monosyllabic words and 15 pairs with disyllabic words. A set of one-way ANOVAs and two ANCOVAs were performed, neither POS nor frequency effects on F_0 were found.

Discussion and conclusion

This study is a preliminary test of the hypothesis that the effect of part of speech on duration is derived from the effect of word frequency, such that the latter is much more robust than the former. The test was performed by comparing pairs of nouns and verbs in Mandarin that are phonetically identical. Results show that there was no significant POS effect on duration or F_0, but there was a significant frequency effect on duration. Given that POS does show a difference in means between nouns and verbs in the same direction as the frequency effect (Tables 1 and 2), it can be concluded that this weak POS effect is derived from a more primitive effect of frequency.

Such frequency effect is consistent with information theory. That is, speakers are under a general pressure to convey as much information as possible in a given amount of time, and this pressure would lead to each word being assigned as little time as possible. But the pressure is balanced by another pressure of communication, i.e., each word also needs to be pronounced as clearly as necessary, so as not to be misheard. But the chance of being misheard can be reduced by the word's predictability, which can be improved by its frequency. So, words that are of higher frequency can afford to have less time, and thus less full articulation. Verbs, as a group, have higher frequency than nouns, and so are likely to be given less articulation time. But ultimately, it is the frequency of each individual word that partially determines its duration. This seems to be supported from the current data. Given the limited size of the dataset in this study, we are making our conclusion cautiously, however. More research in this direction is needed.

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, 31-56.

Aylett, M., Turk, A. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. The Journal of the Acoustical Society of America, 119(5), 3048-3058.

Beijing Language and Culture University. 1986. Modern Chinese Frequency Dictionary.

Coker, C., Umeda, N. and Browman, C. 1973. Automatic synthesis from ordinary English test. IEEE Transactions on Audio and Electroacoustics,21(3), 293-298.

Lightfoot, M. J. 1970. Accent and time in descriptive prosody. Word, 26(1), 47-64.

Names. 2016. [online] Available at: http://www.sosuo.name/tong/.

Shannon, C. E. 2001. A mathematical theory of communication. ACM SIGMOBILE Mobile Computing and Communications Review, 5(1), 3-55.

Xu, Y., Wang, M. 2009. Organizing syllables into groups—Evidence from F0 and duration patterns in Mandarin. Journal of phonetics,37(4), 502-520.