Available online at www.sciencedirect.com

ScienceDirect

Speech Communication 76 (2016) 1–27

SPEECH COMMUNICATION

www.elsevier.com/locate/specom

Predicting tonal realizations in one Chinese dialect from another

Junru Wu [a,b,*], Yiya Chen [b], Vincent J. van Heuven [b,c], Niels O. Schiller [b]

[a] Dept. Chinese Language and Literature, East China Normal University, 500 Dongchuan Rd., Shanghai 200241, China
[b] Leiden University Centre for Linguistics, Leiden Institute for Brain and Cognition, The Netherlands
[c] Dept. Applied Linguistics, University of Pannonia, Egyetem utca 10, Veszprém, Hungary

Received 16 February 2015; received in revised form 20 October 2015; accepted 29 October 2015
Available online 5 November 2015

Abstract

Pronunciation dictionaries are usually expensive and time-consuming to prepare for the computational modeling of human languages, especially when the target language is under-resourced. Northern Chinese dialects are often under-resourced but used by a significant number of speakers. They share the basic sound inventories with Standard Chinese (SC). Also, their words usually share the segmental realizations and logographic written forms with the SC translation equivalents. Hence the pronunciation dictionaries of northern Chinese dialects could be easily available if we were able to predict the tonal realizations of the dialect words from the tonal information of their SC counterparts. This paper applies statistical modeling to investigate the tonal aspect of the related words between a northern dialect, i.e. Jinan Mandarin (JM), and Standard Chinese (SC). Multi-linear regression models were built with between-word pitch distance of JM words as the dependent variable and the following were included as the predictors: SC tonal relations, between-dialect tonal identity, and individual backgrounds. The results showed that tonal relations in SC and between-dialect identity, as predictors featuring the relation between the JM and SC tonal systems, are significant and robust predictors of JM tonal realizations. The speakers' sociolinguistic and cognitive backgrounds, together with the tonal merge and neutral tone information within JM, are important for the prediction of JM tonal realizations and affect the way that between-language predictors take effect.

Keywords: Tone; Translation equivalents; Cognates; Modeling; Individual backgrounds

1. Introduction

1.1. The necessity and sufficiency of modeling under-resourced northern Chinese dialects

Under-resourced languages, featured by the "lack of a unique writing system or stable orthography, limited presence on the web, lack of linguistic expertise, and lack of electronic resources for speech and language processing" (Besacier et al., 2014: 27), have always been a challenge for both engineers of Human Language Technologies (HLT) and linguists. One of the main reasons behind this challenge is the large amount of phonetic data required, which can be both difficult and expensive to acquire. To tackle this challenge, more and more researchers are transferring information from a related language or dialect to improve the understanding and automatic machine-processing of the under-resourced language. For instance, the automatic speech recognition of Afrikaans was significantly improved using the available Dutch data (Imseng et al., 2014). However, to better incorporate the information from the related language, we need a better understanding of the relations between the two languages or dialects. In this aspect, linguists have carried out studies

* Corresponding author at: Dept. Chinese Language and Literature, East China Normal University, 500 Dongchuan Rd., Shanghai 200241, China. Tel.: +86 (0)2154344874.
E-mail address: jrwu@zhwx.ecnu.edu.cn (J. Wu).

http://dx.doi.org/10.1016/j.specom.2015.10.006
0167-6393/

of a wide-range of languages, though linguistic knowledge sometimes needs adaptations to be applied in engineering.

Chinese appears to be anything but under-resourced. For instance, Mandarin Chinese and Shanghai Chinese are already covered by the standardized multilingual text and speech database "GlobalPhone" (Schultz, 2002). Even the (Standard) Mandarin-English bilingual test-to-speech system has seen important breakthroughs (Qian and Soong, 2012). However, compared with the relatively well-investigated Standard Chinese (also referred to as "Mandarin Chinese", "Standard Mandarin", or "putonghua", abbreviated as "SC" in this article), many Chinese dialects are still under-resourced, including most northern dialects.[1] These northern dialects need more attention. First, they are used by a large Chinese population in everyday life (Hamed, 2005; Li, 1988). Second, they are closely related to SC and are often used together with SC. This type of bilingualism comes with frequent code-switching/-mixing and sometimes also results in accented SC speech, which presents challenges for engineers and linguists (Huang et al., 2000; Sproat et al., 2004).

On the other hand, the close relation between the northern dialects and SC is also an attractive resource for the modeling of these dialects. Besides the large overlap in syntactic structure, the northern dialects and SC are very similar in basic sound inventories. For instance, we can find the comparison of the basic sound inventories of major Chinese dialects in a dictionary designed by linguists (Collective_work, 1989). This type of similarity has been proved useful in the sound-to-phoneme modeling in other languages (Imseng et al., 2014; Kamper et al., 2012; Van Heerden et al., 2010). However, there is one additional aspect of the between-dialect relation that may be useful and needs some more exploration. The northern dialects and SC share a high percentage of cognates and frequently borrow from each other[2] (Norman, 2003). The resulting translation equivalents share the same meaning across dialects and sound similar to each other. These related words are easy to identify because they are written in the same characters across all these dialects using the same logographic writing system. This paper applies statistical modeling to explore the tonal aspects of the related words between a northern dialect and SC. As a preliminary but important step before predicting the dialect pronunciation directly from SC pronunciation, the current study investigated to what extent and in what way a very limited but well available SC resource, the SC tonal categories, can predict the dialectal tonal realizations. We also tried to find out how the SC tonal categories, together with the speaker's social and cognitive backgrounds can account for the speaker-dependent tonal variability.

[1] The term "northern dialects" is sometimes distinguished from "Mandarin dialects", which are even more similar to Standard Chinese (Hamed, 2005). Here we use it in a more general way, following Li (1988).

[2] However, the cognates and loan words are difficult to distinguish for closely related dialects.

1.2. Research background on Jinan Mandarin (JM)

We aim at predicting between-word pitch distances for JM Chinese using the tonal relations of the SC counterparts of the target words. JM is a northern dialect of Chinese. It is used in some local TV shows, but mostly in traditional folk arts, such as in "Shandong Kuaishu". Most JM speakers also speak SC fluently, and the mutual intelligibility between JM and SC is high (Tang and van Heuven, 2009). Some linguistic descriptions are available for JM. "Jinan Fangyan Cidian" (JM Dialect Dictionary) (Qian, 1997) provides the largest vocabulary but no recording. "Jinanhua Yindang" (The Sound System of JM Dialect) (Qian and Zhu, 1998) provides recordings of 428 monosyllabic characters, 410 words with two or more syllables, and some sentences. Pronunciations of characters are also available in "Hanyu Fangyin Zihui" (Collective_work, 1989). However, these studies are based on the pronunciations by senior speakers many years ago (above 65 years old in 1993, 1998, and 1979).

Our fieldwork in 2012 showed that JM has become more similar to SC and the differences are mainly only retained in the tonal system. First, the usage and knowledge of JM-specific words are largely reduced and JM-specific words are replaced by words with etymologically related SC counterparts. Second, most JM words are now almost identical to their SC counterparts in segmental structure. However, the tonal differences remain between the JM and SC translation equivalents.

As a result, the current JM dialect shares a high percentage of related words with SC, which are almost only different from their SC counterparts in their tonal realizations (pitch contours). Since most non-tonal resources can already be directly transferred from SC, tone is the main potential space for cost reduction when building the pronunciation dictionary. The building cost of a JM pronunciation dictionary could be reduced if we are able to predict the tonal realizations of the JM words from the tonal information of their SC counterparts.

However, many JM words have shown tolerance of different tonal patterns, possibly due to the on-going process of "lexical diffusion", where new tonal variants have appeared on some words but not on other words originally from the same tonal category (Chen and Wang, 1975; Wang, 1969), and the generalization of JM "neutral tone sandhi" (Qian, 1997), which means some words which were not reported to carry neutral tones are starting to have variants with neutral tone sandhi. As a result, some JM words allow one single tonal pattern (mono-pattern) but the others allow more than one (dual-pattern/multi-pattern). Fig. 1(a) and (b) demonstrate the difference between mono-pattern (i.e. "very", fei1chang2, /feitʂʰaŋ/) and dual-pattern (i.e. "simple", jiandan, /tɕiɛntan/). These words were plotted with normalized F0 contours from multiple speakers. Different tonal patterns of the same word can be observed not only in the production of different speakers but also in the production of the same speaker.