Proc Ling Soc Amer Volume 1, Article 9:1–9
doi:http://dx.doi.org/10.3765/plsa.v1i0.3702

Indefinite markers, grammaticalization, and language contact phenomena in Chinese

Alan Wong*

Abstract. Grammaticalization and language contact are often treated separately, suggesting that these are two distinct, divergent phenomena (Heine & Kuteva, 2003). This however, is known not to be the case. The grammaticalization of an indefinite marker, for instance, can occur due to language contact or be hastened by it (Moravcsik, 2013:200). Contrary to assumptions of many grammarians working in Chinese linguistics, recent work on Standard Chinese (e.g. Chen 2003, Huang 1999) argues that Chinese makes use of determiners. However, few explanations have been given as to why this development has taken place. I suggest that contact with English has encouraged the grammaticalization of the indefinite marker, *yi* 'one' + CLASSIFIER. More specifically, the translation of English texts into Chinese has contributed to the development of an indefinite marker in Chinese (Hsu 1994).

Keywords. grammaticalization; language contact; Chinese; indefinite article

1. Introduction. Previous work on Standard Chinese (henceforth "Chinese") has shown that the construction *yi* "one" + CLASSIFIER (CL) fulfills proposed stages for the cross-linguistic grammaticalization of the numeral 'one' into a marker of indefiniteness (Heine 1997, Chen 2003). Bare nouns in Chinese can be interpreted as definite or indefinite depending on their position relative to a verb (Cheng and Sybesma 1999, 2005). However, by modifying nouns, we can alter how these nouns are interpreted.

(1) a. wǒ xiǎng mǎi shū.
1 want buy book
'I want to buy books.'
b. wǒ xiǎng mǎi yi-běn shū.
1 want buy a-CLASSIFIER book
'I want to buy a book.'

Evidence from a one-million-word balanced Chinese corpus (Lancaster Corpus of Mandarin Chinese) and a translated Chinese written corpus (ZJU Corpus of Translational Chinese) designed as a comparable counterpart suggests that the *yi* + CL construction is indeed continuing to be grammaticalized, and that translation may have a role to play in this development. The translated corpus compared to the original Chinese corpus displayed higher proportions of the generic classifier *ge* and the character for the numeral *yi* "one", despite having fewer nouns overall. Similarly, many other common classifier words were more frequent in the translated corpus.

The many options for expressing definiteness in Chinese often make it possible for translators to successfully capture both word orderings and unambiguously express definiteness of NPs in source texts from languages such as English. Bilinguals, which include translators of texts, are able to draw on resources from more than one language and in doing so break down barriers be-

* For helpful comments and discussion, I would like to thank Raúl Aranovich, Robert Bayley, and Jack Hawkins, as well as my fellow graduate students at UC Davis. Additionally, I would like to thank the anonymous reviewers of the LSA abstracts and those that attended my presentation on this topic for their valuable feedback. Finally, I wish to thank my native Chinese speaker friends for their invaluable tacit linguistic knowledge.
Author: Alan Wong, University of California Davis (alnwong@ucdavis.edu).

tween languages via conceptual transfer (Matras 2011, 2010, 2007). This small corpus study suggests that translation may motivate the use of potentially 'marked' morphosyntactic forms. These forms, though originating from writing, may become increasingly incorporated into the speakers' repertoires as individuals interact with translated texts.

2. Chinese. Modern Standard Chinese, oftentimes referred to as 'Mandarin' or simply just 'Chinese' is the official language of the People's Republic of China (PRC), where it is referred to as *Pǔtōnghuà* "Standard Speech" and in Taiwan, where it is referred to as *Guóyǔ* "national language". The standard pronunciation and grammar of Chinese is associated with the Beijing region of north China, though not Beijing itself (Yip and Rimmington 2006, Zhang 2005). This paper is primarily concerned with the written form of this language, though it is assumed that writing may in turn eventually influence speech.

Chinese is a relatively isolating language, which is relatively impoverished in terms of inflectional morphology. However, unlike English, Chinese is a topic-prominent language which consistently presents topics (known information; what sentences are 'about') in sentence initial position, followed by a 'comment', or 'predicate', which says something about the topic (Chao 1968, Li and Thompson 1989, Wiedenhof, 2015). The semantics of pre-verbal and post-verbal objects are quite varied, as has been noted by Chao (1968) and others. Consequently, it is difficult to classify Chinese in terms of the basic constituents S, V, and O. Chinese, nevertheless, is often classified as an 'SVO' language like English because the unmarked way of presenting transitive sentences in many situations conforms to this word order.

Modern Standard Chinese, like other standard language varieties in East Asia, arose amidst nationalism and the formation of a modern nation-state. In the late nineteenth century, the concept of a national language was introduced into Chinese discourse from Japan (Chen 2007:145). Prior to this period, efforts to standardize language in China were primarily directed at written forms, specifying proper shapes and pronunciations of Chinese characters. Through the National Language movement of the early twentieth century, efforts were made to establish and promote a standard language throughout China in both spoken and written form (Chen 2007:145). Efforts were made to shift from the written stand based on Old Chinese to a variety closer to the contemporary vernacular.

Though Chinese is based on Beijing Mandarin, particularly in vocabulary and grammar, it is not the native language of any real set of individuals, as it is essentially the spoken form of the vernacular literary language used by contemporary Chinese writers. Generally, Chinese excludes local expressions particular to specific topolects, and incorporates words, phrases, and grammatical forms from other Chinese dialects, Old Chinese, and foreign languages (Chen, 2007:146).

The interaction of English and Chinese is particularly interesting for a number of reasons. Firstly, Chinese and English share many typological features. For example, both can be regarded as VO basic constituent ordering types. Second, Chinese and English are historically very distant. This fact acts as a sort of control ensuring that similarities and differences between Chinese and English are not specific products of shared heritage, but result from more general linguistic properties. Thirdly, Chinese has a widely used written language, which is convenient for carrying out corpus studies.

In China, as in many other places in the world, English is regarded as prestigious. English is a required subject in school, so at least some degree of exposure to English is expected for all liter- ate Chinese speakers. Nevertheless, the argument presented here, that Chinese is being impacted by translation, does not hinge on whether or not high degrees of Chinese/English bilingualism are present. Rather, what is assumed is that translations widely read, at least by rela-

tively well-educated, literate people, and that exposure to these translations may alter the frequencies in which Chinese speakers encounter particular English-influenced forms, regardless of whether or not they are conscious of this influence.

2.1. EXPRESSING DEFINITENESS IN CHINESE. Nouns in Chinese have a single, invariant form and lack marking reflecting definiteness, number, case, etc. (Ross and Ma, 2006). This contrasts with English nouns, which express grammatical number, for example. Definiteness is expressed in numerous ways in Chinese. The following means of expressing definiteness will be discussed:

- Word order
- Lexical semantics
- Morphology
- Syntactic frames
- Definiteness markers

Unlike English, Chinese does not usually employ overt markers (articles) to encode definiteness. The most common way of expressing definiteness in Chinese through word order and context in discourse. Word order typically conveys information on definiteness in Chinese because known referents are typically definite and given first, while new referents are introduced second, often as indefinite referents:

(3) Hufei mǎi shū qù-le.
Hufei buy book go-PERFECTIVE/CURRENTLY RELEVANT STATE
'Hufei went to buy book(s).' (Indefinite)

(4) Hufei hē-wán-le tang.
Hufei drink-finished-PERFECTIVE soup
'Hufei finished the soup.' (Definite)

These examples are 'typical' in that known referents are presented first, if possible. Then, the VO word order is followed. Oftentimes, as in these examples, the 'unmarked' VO word order in Chinese is adequate to convey expected definiteness encodings. Pronouns, which are definite, often occur in subject position. New objects may be introduced into the discourse as indefinite by being presented after the subject.

This principle of old information first, new information second is exemplified by the presentative construction. The presentative construction is used for introducing new information, which most often corresponds with indefinite reference. In Chinese, subjects are placed at the beginning of the sentence, followed a predicate (or a "comment" under some analyses). Oftentimes, this means that the second part of a sentence presents *new information*. In presentative sentences, an indefinite noun phrase is "presented" into discourse. Such sentences lack topics and need not be about a definite or generic noun phrase.

(5) (Li and Thompson 1989:91)
a. yǒu rén zài dǎ diànhuà gěi Zhāngsān.
exist person DURATIVE hit telephone to Zhangsan
'Someone is making a call to Zhangsan.'
b. jìn lái le yi-ge rén.
enter come PERFECTIVE a-CLASSIFIER person
'A person came in.'

Similar patterns are found in topicalization constructions.

All languages have certain words and expressions that inherently convey information on definiteness. Individual lexical items may encode definiteness in their semantics. Deictic expressions, for example, point to things available in the given context. Thus, *zhèige* 'this' and *nèige* 'that' inherently have definite reference because of their deictic function. On the other hand, in Chinese, interrogative pronouns are used for indefinite reference (Wiedenhof 2015:215).

(6) (Wiedenhof 2015:215)

a. tāmen gēn shéi qù.
they with who go
'They are going with just anyone.'

b. tā mǎi shéme Yīngwén shū.
he buy what English book
'He is buying this or that English book.'

In the sentences above, WH-words (*shéi* 'who' and *shéme* 'what') are used to make indefinite reference. Semantically, this makes sense, because a *question* word is used to refer to an entity that cannot be identified. As Mandarin does not make use of changed word order to express questions as English does, these two sentences, uttered with different intonation would be interpreted as "Who are they going with?" and "What kind of English books is he buying?", respectively.

Morphological processes may imply certain interpretations of definiteness. The "collective suffix" *-men*, which is attached to nouns for animate entities (which are nearly always humans) necessarily expresses definiteness (Wiedenhof 2015:302). A bare noun, such as *péngyou* 'friend(s)' may have either definite or indefinite reference. But, suffixed with *-men*, the word *péngyou-men* always refers to a discourse identifiable (that is, definite) collection of *péngyou* 'friend(s)'.

(7) a. tā de péngyou bu duō.
3 SUBORNINATION friend not much
'He does not have many friends.'

b. * tā de péngyou-men bu duō.
3 SUBORNINATION friend-COLLECTIVE not much
'He does not have many friends.'

In the examples above, the second sentence is not preferred because the friends mentioned do not act as a group. The bare noun form *péngyou* in the first sentence appropriately represents friends as neither singular, plural, nor a group. Whereas in the second sentence, suffixing *-men* to *péngyou* attempts to force a definite interpretation, which is awkward or ungrammatical.

Just as morphological processes, such as adding the collective suffix *-men* to NPs, leads to certain interpretations of definiteness, there are syntactic frames which demand particular interpretations. For example, the *bǎ* construction typically requires definite reference for the direct object following *bǎ* (Wiedenhof, 2015, 155). The general structural frame for this form is:

(8) Li and Thompson (1989:463)

SUBJECT *bǎ* DIRECT OBJECT VERB

Typically, Chinese does not mark objects with any overt morphology. In this way, the *bǎ* construction is marked because the direct object is marked with *bǎ*. Wiedenhof (2015:155) states

that the *bǎ* construction contributes two meanings: (a) definiteness, and (b) impact on the object. Li and Thompson (1989:466) likewise state that this construction usually (though not always) denotes definite reference, and that it additionally conveys a notion of "disposal", by which something *hap- pens* to the direct object. This construction, and others in Chinese, convey specific interpretations of definiteness.

While the interpretation of definiteness is often unambiguous as a consequence of the semantics of some construction, or the lexical semantics of some item (such as with the deictic words and question words), many contexts permit either a definite or indefinite interpretation of some NP. Chinese, has resources for explicitly coding definiteness in these otherwise definiteness unspecified situations via the use of definiteness markers. The focus of the corpus study portion of this paper is the increased use of definiteness markers in Chinese. In (1a) below, given in the introduction of this paper, the first sentence does not mark *shū* 'book' for any particular definiteness value. If *shū* 'book' is already established in discourse, it may be interpreted as having definite reference alongside the subject of this sentence, *wǒ* 'I'. On the other hand, if there is no item already established in discourse that *shū* 'book' might refer to, its reference will be determined as indefinite.

(9) a. wǒ xiǎng mǎi shū.
1 want buy book
'I want to buy books.'
b. wǒ xiǎng mǎi yi-běn shū.
1 want buy a-CLASSIFIER book
'I want to buy a book.'

On the other hand, in (1b), *yi-běn* 'a + CL' marks the noun *shū* 'book' as indefinite. Unlike in the example above, this example does not permit a definite interpretation under normal circumstances. Thus, functionally, definiteness markers in Chinese are able to explicitly mark an N or NP for definiteness where word order may permit multiple interpretations. For pragmatic reasons, a certain word order which usually results in a certain interpretation of definiteness may be preferred over another. Definiteness markers may be used to 'override' these expectations.

3. Corpus Study. Through the investigation of counts of the structure *yi* + CL across two corpora, this study aimed to establish that translation, a site of language contact, could contribute to the productive use of indefinite markers in Chinese.

3.1 GRAMMATICALIZATION OF THE INDEFINITE ARTICLE. It has been argued that the genesis of articles (both indefinite and definite) has been triggered by the influence of neighboring languages (Heine and Kuteva, 2003). While this sort of influence can be called a 'language contact phenomenon' it differs from typical borrowing in that phonological matter is not borrowed, but rather a morphosyntactic idea is borrowed.

As described in the previous section, indefinite reference can be expressed in Chinese explicitly (rather than implicitly by predicate interpretation) via the use of the structure *yi* + CL. Chinese then, has the syntactic resources to encode indefiniteness of some NP, in many sentence positions, as the indefinite article can do in languages like English. Chen (2003) calls this construction the "indefinite determiner" and argues that it fulfills also the stages of grammaticalization from a numeral to a generalized indefinite determiner.

Indefinite articles have been observed cross-linguistically to arise from the numeral "one". In English, for example, as is also the case with Chinese, I will argue, the indefinite marker *a/an* evolved from an unstressed form of the numeral one (Bybee 2015:79). Below is the path of

grammaticalization from the numeral one to an indefinite marker, as summarized by Moravcsik (2013:200):[1]

- **Initial Stages**: Given an indefinite article, it is likely to have arisen from the numeral *one*.
- **Final Stages**: Given the numeral *one*, it may change into an indefinite article.
- **Intermediate Stages**: Both changes are instances of grammaticalization – a gradual phonological, semantic, and if applicable, morphological reduction.
- **Conditions**: Language contact may trigger or accelerate the development of articles.

This paper proceeds from the assumptions that *yi* + CL already fulfills the "Initial Stages" and "Final Stages" listed by Moravcsik (2013). This form clearly derives from a construction involving the numeral 'one'. Likewise, this form is able to be used as an article (albeit not one that is required for NPs in nearly as many contexts as in English, for example). What remains to be explained then, are the "Intermediate Stages" and the "Conditions" that accompanied this extension of the use of *yi* + CL. By Moravcsik (2013)'s criteria, we must show that both phonological and semantic, and optionally morphological, reduction is underway.

The indefinite marker yi + CL displays phonological reduction because it can be contrasted with yí + CL (note the diacritic indicating a second time on *yí*). While *yi* + CL (with a neutral tone, indicated by a lack of any diacritic) is interpreted as a marker of indefiniteness, a stressed *yí* carries the meaning of the numeral one (Wiedenhof 2015:253). This situation parallels the distinction between sentences such as, 'I would like *a* piece of pizza' as opposed to 'I would like a piece of pizza', where the emphasis on *a* in the first sentence demands an interpretation of *a* meaning 'one', whereas in the second, *a* left unstressed is simply an indefinite article.

Orthographically, *yí/yi* both used as a numeral and as an indefinite marker employ the same Chinese character, 一. The fact that these two phonetically and semantically distinct usages share the characters highlights the association between *yī/yì* used as a numeral and as an indefinite marker. A shared orthographic form combined with speakers' suggests that what once may have been thought of as one single lexical item has diverged into multiple forms (i.e. the original lexical item *yī* 'one' and the newer grammaticalized item *yi* (+ CL) marking indefiniteness). The form *yi* + CL can be said to semantically bleached of its older meaning of 'one', which is expressed with the now contrasting form *yí* + CL.

Continuing from Moravcsik (2013)'s criteria, what now remains to be explained is the *conditions* for language change, of which Moravcsik (2013) writes, "language contact may trigger or accelerate the development of articles". The remainder of this paper will argue from a small quantitative corpus study why we have good reason to believe that this is the case for Chinese.

3.2 DATA. For this study, two corpuses were used: The Lancaster Corpus of Mandarin Chinese version 1 (LCMCv1, Brown family, 1991) and the ZCTC corpus (ZJU Corpus of Translational Chinese), both available freely online. Both corpora are balanced 1 million word corpora, and the ZCTC (henceforth the "translational corpus") was explicitly designed as a parallel counterpart to the LCMCv1 (the "native corpus"). While these corpora are rather small in size, they offer an opportunity to study differences between translated and native text not currently available using larger corpora in Chinese and English. Both of these corpora are balanced across a number of genres, such as press reportage, press editorial, biography and essay, science, general fiction, and humor.

[1] Moravcsik (2013) also lists stages for the definite article, which I have omitted in this paper.

3.2 RESULTS. Forms corresponding to the indefinite marker were counted across the two corpora. It is hypothesized that greater use of the indefinite marker corresponds to further grammaticalization of this form. The translated corpus represents a site of more intense language contact than does the native corpus. My argument then, is that if we observe greater usage of the innovative form (the indefinite marker) in the translated corpus, we can attribute these differences in frequency to language contact phenomena. This will fulfill Moravcsik's (2013) *conditions* of grammaticalization, which says "language contact may trigger or accelerate the development of articles".

Item	Native Corpus	Translational Corpus
yi+ *ge* ("a/an")	2832	3782
yi + *xie* ("some")	695	995
yi ("one") +	8365	8677
Total:	11,892	13,454

$t(2) = 99.54, p < 0.05$

Table 1: Indefinite Marker Structure Counts

Table 1 lists counts of indefinite structures in the two corpora. In the item in the first row, *yi* is the Chinese numeral 'one', and *ge* is the generic classifier. Together, *yi* and *ge* make up a fully functionally indefinite article, which previous literature has described. The item in the second row, combines *yi* 'one' with *xie* 'few', to give a plural indefinite marker, comparable to 'some' in English, or the determiner "*unos/unas*" in a language like Spanish. Finally, the item in the third row, the numeral one, *yi*, alone is always paired with classifier words to form *yi* + CL compounds, such as *yi* + *ge* (generic) in the first row. *Yi* + *ge* (first item) was included in this table as a separate entry from the third item because it was listed as its own word in the corpora I looked at and because it is expected that as this form continues to grammaticalize, it is this form with the generic classifier that will come to dominate the role of indefinite marker.

Item	Native Corpus	Translational Corpus
Nouns	180,195	171,652
Indefinite Structures	11,892	13,454
Nouns – Indefinite Structures	168,303	158,198
Nouns Marked as Indefinite	6.60%	7.84%

$t(1) = 201.51, p < 0.05$

Table 2: Noun Counts

Running a Chi-squared test of independence on the counts of nouns without indefinite structures (3rd row) and nouns with one of the indefinite structures (we will assume that the indefinite structures are always followed by nouns), we get the results listed in Table 2. Counts of non-generic classifiers are listed in Table 3. Unlike the items listed in Tables 1 and 2, the items in this table do not pattern in a clear-cut way. While for most items, the translational corpus has higher counts, this is not always the case. Namely, *shǒu* and *zhāng* are more numerous in the native corpus, and other counts are quite close.

Item	Native Corpus (LCMCv1)	Translational Corpus (ZCTC)
本 *běn*	77	152
份 *fèn*	101	273
辆 *liàng*	88	143
首 *shǒu*	34	32
双 *shuāng*	65	81
条 *tiáo*	708	751
位 *wèi*	804	988
张 *zhāng*	344	236
只 *zhī*	228	296
$t(8) = 121.88, p < 0.05$		

Table 3: (Non-generic) Classifier Counts

4. Discussion. In the proceeding analysis, the following assumptions will be made: (1) Both corpora are taken to be representative of the written language, at least, of Chinese, (2) the translated corpus has relatively more influence from English than the native corpus, (3) text is accessible to speakers of Chinese (i.e. the population is literate) and may influence speakers' mental representations of language.

The first of these assumptions captures a basic idea of studying language using corpora. A corpus is a sample of linguistic data. All things being equal, a larger corpus is better better than a smaller corpus because a bigger sample size is more likely to accurately reflect actual characteristics of a population. Nevertheless, *balanced corpora*, corpora that have been compiled in such a way to be more representative of a language in some way are a valuable resource because they attempt to correct for some difficulties that may arise from indiscriminately analyzing large quantities of data. Naturally, there are certain topics that are more frequently written about than others. Being more frequently written about, however, does not necessarily mean that these topics are more representative of general language use — an especially prolific or verbose subset of the population may be overrepresented without attention to balance. Likewise, some of topics frequently discussed in writing may be scarcely discussed in spoken language. The hope of using balanced corpora then, is to get a better balanced sample, which may also be potentially more informative about spoken language.

The second assumption is made to address the fact that grammaticalization often occurs language- internally, without the influence of outside languages. Changes in Chinese, or any other language for that matter, cannot be attributed to language contact situation simply because language contact is present. In this way, the native corpus serves as a sort of "control" group. While both the native and translated corpora are subject to forces of grammaticalization, the translated is *especially* subject to forces of language contact in ways that the native corpus is not.

The third assumption is needed for any argument that texts actually influence speakers. We must assume that our population of interest (Chinese speakers) do read texts, of which we hope our corpora are representative of. As explained in the first half of this paper, we have many reasons to believe that text can influence languages in speakers. By reinforcing certain patterns of usage (including introducing new words), text 'trains' speakers to use language in a particular way the same way spoken language does. Psycholinguistic evidence supports the idea that speakers have rich, dynamic representations of language in the mind, which change over time.

5. Conclusion. The significant differences between the native and translated corpora suggest that Moravcsik (2013)'s *conditions* for the grammaticalization of the indefinite article have been ful-

filled in Chinese. As the translated text quantitatively different from native text, there is reason to believe that translation, as a site of language contact, may be involved in grammaticalization processes. As language change can occur gradually, with changes in frequencies accumulating over time leading to larger results, this corpus study provides evidence of English accelerating the grammaticalization of the indefinite article in Chinese.

References

Bybee, Joan. 2015. *Language Change*. Cambridge: Cambridge University Press.

Chao, Yuan Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley, CA: University of California Press.

Chen, Ping. 2003. Indefinite determiner introducing definite referent: a special use of 'yi 'one' + classifier' in Chinese. *Lingua* 113. 1169-1184.

Chen, Ping. 2007. China. In Andrew Simpson (eds.), *Language and Society in East Asia*. New York: Oxford University Press.

Cheng, Lisa Lai-Shen & Sybesma, Rint. 1999. Bare and Not-So-Bare Nouns and the Structure of NP. *Linguistic Inquiry* 30(4).

Cheng, Lisa Lai-Shen & Sybesma, Rint. 2005. Classifiers in Four Varieties of Chinese. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford Handbook of Comparative Syntax*. New York: Oxford University Press.

Heine, Bernd. 1997. *Cognitive Foundations of Grammar*. Oxford: Oxford University Press.

Heine, Bernd & Kuteva, Tania. 2003. On contact-induced grammaticalization. *Studies in Language* 27(3). 529-572.

Hsu, Jialing. 1994. Englishization and Language Change in Modern Chinese in Taiwan. *World Englishes* 13(2). 167-184.

Huang, Shuanfan. 1999. The Emergence of a Grammatical Category Definite Article in Spoken Chinese. *Journal of Pragmatics* (31). 77-94.

Li, Charles N. & Thompson, Sandra A. 1989. *Mandarin Chinese: A Functional Reference Grammar.* Berkeley, CA: University of California Press.

Matras, Yaron. 2007. The Borrowability of Structural Categories. *Approaches to Language Typology* 38(31).

Matras, Yaron. 2010. Contact, Convergence, and Typology. In Raymond Hickey (ed.) *Handbook of Language Contact*. Oxford: Blackwell. 66-85.

Matras, Yaron. 2011. Grammaticalisation and Language Contact. In Heiko Narrog & Bernd Heine (eds.) *The Oxford Handbook of Grammaticalisation*. 279-290.

Moravcsik, Edith A. 2013. Language in Flux. In *Introducing Language Typology.*

Ross, Claudia & Ma, Jingheng Sheng. 2006. *Modern Mandarin Chinese Grammar.* New York: Routledge.

Wiedenhof, Jeroen. 2015. *A Grammar of Mandarin*. Philadelphia: John Benjamins Company.

Yip, Po-Ching & Rimmington, Don. 2006. *Chinese: An Essential Grammar*. New York: Routledge.

Zhang, Qing. 2005. A Chinese Yuppie in Beijing: Phonological Variation and the Construction of a New Professional Identity. *Language in Society* 34. 431-466.