Stylistic variation in L1 and L2 Chinese
Native speakers, learners, teachers, and textbooks

Xiaoshi Li
Michigan State University

This study examines stylistic variation patterns in L1 and L2 Chinese, focusing on two linguistic structures: morphosyntactic particle DE and subject pronoun. The data were from thirteen native speakers, four Chinese instructors, twenty-three L2 Chinese learners, and four Chinese textbooks. Results from variation analysis with frequency description show four general patterns. First, instructors used overt forms of stylistic variants in class significantly more frequently than native speakers did in conversations. Second, learners tended to overuse the overt forms compared with their native speaker peers. Third, learner patterns of stylistic variation aligned closely with those of their teachers. Finally, unlike teacher input, textbook input demonstrated mixed results compared with learner patterns. For DE use, learners’ patterns aligned significantly with those in textbooks, but did not for subject pronoun use. The implications for stylistic variation in conversational and written Chinese are discussed and suggestions are offered for Chinese and foreign language instruction.

Keywords: stylistic variation, Chinese, sociolinguistics, foreign language instruction

1. Introduction

Native speakers frequently use a variety of language forms to express similar meanings. In addition, depending on different linguistic and social contexts, these forms might carry varying stylistic meanings in terms of formality, clarity, tone, and politeness, among other qualities. Language learners often struggle with using the appropriate language variations in authentic social contexts even after years of language learning. Learners need to acquire not only the correct language forms but also, more importantly, the nuances of the stylistic meanings that these forms convey, and be able to produce variable speech styles appropriately. To this end,
an understanding of native speakers’ and learners’ patterns of language variation, how and why their patterns are different, and the factors influencing the patterns, is necessary.

Previous studies on sociolinguistic variation have found a variety of reasons for the variability in learners’ interlanguage, including linguistic, social, psychological, developmental, and educational factors (Bayley, 1994, 1996; Dewaele & Mougeon, 2004; Dewaele, 2004; X. Li, 2010a; Mougeon & Rehner, 2001; Mougeon, Rehner, & Nadasdi, 2004; Rehner, Mougeon, & Nadasdi, 2003; Wolfram, 1985; Young, 1991). In recent decades, there have been a number of variation studies, especially in French, that showed empirically that teachers and textbooks have a significant impact on learners’ use of sociolinguistic variants. Mougeon and his associates, through investigation of a variety of French language structures, found that vernacular speech is not authentically represented by teachers and textbooks, which overuse the prestige forms of variants. The researchers claim this overuse to be a “disservice” to the students (Dewaele & Mougeon, 2004; Mougeon, Nadasdi, & Rehner, 2010; Mougeon & Rehner, 2001; Mougeon, Rehner, & Nadasdi, 2004; Nadasdi, Mougeon, & Rehner, 2005, 2008; Rehner & Mougeon, 1999, 2003; Rehner, Mougeon, & Nadasdi, 2003). They also found that educational factors, including teachers’ input and instructional materials, have a significant impact on the use of sociolinguistic variants by French L2 or immersion learners, and that students follow teachers’ patterns of sociolinguistic use of variants. For example, Etienne and Sax (2009) examined 22 French textbooks in the U.S. and found that the three target stylistic variants (i.e. on vs. nous, ne deletion vs. ne retention, and interrogative structures) were not represented consistently and their authentic uses were often misrepresented. Therefore, the researchers called for the redesign of the French language textbooks to incorporate more authentic use of the language.

Corresponding to the increasing research interest in sociolinguistic variation, more and more SLA researchers have become interested in the implications of sociolinguistic research for foreign language teaching: whether it is necessary to explicitly teach sociolinguistic competence in the classroom, and if so, how (Cohen, 2008; Etienne & Sax, 2006; 2009; Lemmerich, 2010; Lyster, 1994; Taguchi, 2015; Valdman, 2003). For example, De Vito (1991) advocated incorporating native speaker norms into French immersion and foreign language education based on the mismatch found in her study between French native speaker use of four grammatical features and the treatment of these features in French second language textbooks. In another study, Harlow (1990), examining the differences of speech act strategies used by French native and nonnative speakers, advocated the incorporation of social variables such as age and familiarity between speakers into the teaching of speech acts in French. Valdman (2003) proposed a pedagogical norm that takes into consideration the fact that members of any linguistic
community shift their speech styles depending on their social contexts and intentions. She found that explicit instruction of variable forms resulted in learners’ “more accurate performance and less puristic attitudes toward language variation” (2003, p. 76).

This study aims to contribute to this line of research in the context of L2 Mandarin Chinese (hereafter Chinese). The study examines the stylistic variations demonstrated in the speech of L2 Chinese (CSL) learners, Chinese native speakers, and Chinese teachers, as well as in advanced-level L2 Chinese textbooks. Research questions include:

What are the general patterns of the use of stylistic variants by L2 Chinese learners, Chinese native speakers, Chinese teachers, and Chinese language textbooks? Are they different? If so, how?

Before I present my study, in the following sections, I will first present the definition of sociolinguistic competence and stylistic variation, followed by descriptions of the target language features.

2. Sociolinguistic competence and stylistic variation

Since Hymes’s proposal of the concept of communicative competence, which emphasized the importance of “what to say to whom in what circumstances and how to say it” (Hymes, 1972, p. 277), scholars and language instructors have approached and explored this concept from different perspectives. One perspective involves conceptualizing appropriateness of language use in a social context (Dewaele, 2008). In the field of sociolinguistics, Lyster (1994, p. 263) defined the concept of sociolinguistic competence as “the capacity to recognize and produce socially appropriate speech in context.” An important component of sociolinguistic competence is the ability to style-shift appropriately. For example, ne in French, the first particle of negation, is a widely-studied sociolinguistic variant. L2 French learners need to know when to omit it, because keeping it all the time would sound too formal or like classroom French. Consequently, researchers have conducted an array of studies investigating relevant factors that affect the use of target linguistic forms (Abreu, 2009; Lyster, 1994; Mougeon & Rehner, 2001; Mougeon, Rehner, & Nadasdi, 2004; Rehner, Mougeon, & Nadasdi, 2003). Besides linguistic factors relevant to the target variants that influence learners’ use of the variants, studies have found that social, developmental, and educational factors, such as proficiency level, gender, age, social network, interlocutor, time spent in the target language environment, instructor speech, and textbooks, also play very important roles. This study investigates L2 Chinese learners’ stylistic variation patterns and compares
them with variation patterns demonstrated by Chinese native speakers, learners’ instructors, and textbooks, aiming to explore the role of instructional factors in L2 Chinese stylistic variation. The study focuses on variation in the use of two linguistic forms: morphosyntactic particle DE and subject pronoun in Chinese.

3. Morphosyntactic particle DE and subject pronoun use in Chinese

Both the morphosyntactic particle DE and subject pronoun are sociolinguistic variants in Chinese and their uses demonstrate variation by native speakers (retention vs. deletion of DE; overt vs. null form for subject pronoun). Their use is influenced by various linguistic and social factors (Chang, 1994; Chappell & Thompson, 1992; Jia & Bayley, 2002; X. Li, 2010b; X. Li, Chen, & Chen, 2012; Liu, 2003). In this section, I briefly introduce the relevant literature on these two variants and their linguistic functions and environments.

3.1 Morphosyntactic particle DE

DE is the most frequently-used word in Chinese (Xiao, Rayson, & McEnery, 2009) and features optionality and multifunctionality (Shi & Li, 2002). DE has three primary grammatical functions: genitive marker, attributive marker, and nominalization marker (Li & Thompson, 1981; Yip & Rimmington, 2004). Based on empirical data, X. Li (2010a, 2010b) ascertained that DE use is optional in these three major functions. DE use is also optional when DE structures are immediately followed by a demonstrative (D) or a number+classifier (NC) structure. This study focuses on optional cases of DE, and the following is a brief explanation of each function with examples extracted from the L2 learner dataset.

As a genitive marker, DE shows the possessive relationship between the modifier and the head noun (N) or noun phrase (NP), as in Example (1).

1. Transcription conventions Li and Thompson (1981):

<table>
<thead>
<tr>
<th>ATT</th>
<th>ATTRIBUTIVE (de)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL</td>
<td>CLASSIFIER</td>
</tr>
<tr>
<td>COP</td>
<td>COPULAR</td>
</tr>
<tr>
<td>CRS</td>
<td>CURRENT RELEVANT STATE (le)</td>
</tr>
<tr>
<td>DEM</td>
<td>DEMONSTRATIVE</td>
</tr>
<tr>
<td>GEN</td>
<td>GENITIVE (-de)</td>
</tr>
<tr>
<td>NEG</td>
<td>NEGATION</td>
</tr>
<tr>
<td>NOM</td>
<td>NOMINALIZER (de)</td>
</tr>
<tr>
<td>PFV</td>
<td>PERFECTIVE ASPECT (-le)</td>
</tr>
<tr>
<td>PL</td>
<td>PLURAL</td>
</tr>
<tr>
<td>PREP</td>
<td>PREPOSITION (zài).</td>
</tr>
</tbody>
</table>
(1) 他们 的 父母
    tā-men DE fùmù
    he-pl gen parents
    ‘their parents’

As an attributive marker, DE can connect the head N/NP with an adjective (ADJ DE N/NP), a noun (N DE N/NP), a verb (V DE N/NP), a phrase (Phrase DE N/NP), or a relative clause (Subject DE N/NP). See Examples (2)–(6).

(2) 我 得 知 道 很 多 新 的 情 况。
    wǒ dĕi zhīdào hĕnduō xīn DE qíngkuàng
    I have to know many new ATT situations
    ‘I have to get to know many new situations.’

(3) 俄罗斯 的 城市
    éluósi DE chéngshì
    Russia ATT city
    ‘Russian cities’

(4) 喝 的 水 不 好
    hē DE shuǐ NEG hăo
    drink ATT water not good
    ‘The water (I) drank was not good.’

(5) 吃 饺 子 的 时 候 也 很 多。
    chī jiăozi DE shíhou yĕ hĕn duō
    eat dumplings ATT times also very many
    ‘There were many times (I) ate dumplings.’

(6) 我 最 喜 欢 的 城 市 是 上 海。
    wŏ zuì xĭhuān DE chéngshì shì shànghăi
    I most like ATT city cop Shanghai
    ‘My favorite city is Shanghai.’

As a nominalization marker, DE occurs when there is no head noun or noun phrase (N/NP) and helps the modifier acquire a nominal feature. There are also cases when DE does not nominalize the modifier. I include them in this category because they share the feature of having no head word. See Example (7), where DE nominalizes the modifier 喜欢 (xĭhuān, 'like').

(7) 我 喜 欢 的 是 游 泳。
    wŏ xĭhuān DE shì yóuyŏng
    I like NOM COP swim
    ‘What I like is swimming.’
When DE is followed by a demonstrative (D) or number (N)+classifier (C), it is often omitted, as in Example (8).

(8) 我 (的) 那 个 朋友
wǒ (de) nà gè péngyou
I (GEN) DEM CL friend
‘That friend of mine’

Studies have shown that DE use by Chinese native speakers is influenced by various factors. Chappell and Thompson (1992) and Liu (2003) examined the use of DE in a phrasal structure, NP1 DE NP2, and found that DE use by native speakers is influenced by a combination of pragmatic, structural, and semantic factors. Chang (1994) explored DE use in another phrasal structure, Adj. DE NP, and showed that DE use is influenced by the information status and discourse functions of the NP, and the relationship between the NP and other possible candidate NPs within a shared discourse. X. Li (2010b) included every function of DE in her investigation and explored both linguistic and social factors. Her results showed that DE use is influenced by DE functions as well as age, occupation, gender, and formality. Females, teachers, and older speakers tend to use DE more than males, students, and younger speakers. DE is used more in formal than in informal settings. Greater use of DE sounds more formal and literary. As to DE use by L2 Chinese learners, X. Li (2010a) showed that it is influenced by DE functions, proficiency level, length of residence in China, and gender. Higher proficiency level, longer time in a target-language environment, and being female all tend to promote DE deletion. In addition, learners’ patterns align with the patterns in their teachers’ speech and textbooks.

3.2 Subject pronoun use

In Chinese, both subjects and objects can be realized through a full NP, a personal pronoun, or a null form. This study focuses on the alternation between personal pronoun and null form in the subject position as illustrated in Example (9). As shown here, the first mention of the subject (我 meaning ‘I’) is realized by a personal pronoun and the second mention is in the null form.

(9) 我 在 阳 朔 的 时 候, Ø 认 识 了 很 多 很 多 旅 游 者。
wǒ zài yánghuò shíhou, Ø rènshi le hěnduō hěnduō lǚyóuzhě
I PREP Yangshuo ATT time, Ø know PFV very many very many tourists
‘When I was in Yangshuo, I got to know many tourists.’

It is widely acknowledged that one of the major constraints on subject pronoun use involves referent switch: whether or not there is a switch between subject
referents in two adjacent clauses (Bayley & Pease-Alvarez, 1996, 1997; Cameron, 1992, 1993; Flores-Ferrán, 2002, 2004; Jia & Bayley, 2002). There are two major conditions among referents in adjacent clauses. The first is when the subject referents are the same, as in Example (10).

(10) 我有时候 晚上 回家， Ø 看 电视 后 Ø 就 睡觉。
    wǒ yǒushíhòu wǎnshàng huíjiā, Ø kàn diànshì hòu Ø jiù shuìjiào
    I sometimes evening go home, Ø watch TV after Ø then go to bed
    ‘Sometimes I go back home in the evening. I will go to bed after watching
    TV.’

The other condition is when the subject referent changes in the second clause, as in Example (11).

(11) 我 来哈尔滨以后， 他走了。
    wǒ lái hāĕrbīn yĭhòu, tā zŏu le
    I come Harbin after, he leave PFV
    ‘After I came to Harbin, he left.’

When there is a referent switch between two adjacent subjects, several cases need further categorization. First, the subject pronouns in the two clauses could be different, but the underlying entities are the same, as in Example (12). In this example, the underlying referents of the omitted subject “you” in speaker A’s question and the subject “I” in speaker B’s answer are actually the same person.

(12) A. 八月份 的时候 Ø 做 什么？
    bā yuèfèn de shíhou Ø zuò shénme
    August ATT time Ø do what
    ‘What are you going to do in August?’

    B. 我 还 没 决定。
    wǒ hái méi juédìng
    I still not decide
    ‘I haven’t decided yet.’

In another case of a referent switch, the subject referents could partially overlap between two adjacent clauses. See Example (13).

(13) 他是 杨玉林 的 朋友。 他们 一起 当兵。
    tā shì yáng yùlín de péngyŏu. Tā-men yīqĭ dāngbīng
    He COP Yang Yulin GEN friend. He-pl together join:army
    ‘He is Yang Yulin’s friend. They joined the army together.’

In the third condition, the subject referent in the preceding clause is not in subject position, but in another part of the sentence (W. Li, 2004, 2006); this is possible
because of that unique and frequent feature in Chinese language, topic chain. The subject can be an object in the preceding clause (Example 14), a fronted topic (Example 15), or in another position in the sentence structure (Example 16).

(14) 实际上我认识她。她住在商业大学。
qìshí wǒ rènshí tā. Tā zhù zài shāngyè dàxué
‘I actually know her. She lives at the University of Commerce.’

(15) 这点，我觉得 Ø 不太好。
zhèdiăn, wǒ juéde Ø bù tài hǎo
‘I think this is not very good.’

(16) 他们很长时间跟我争斗。然后，我长大了，
tā-men hěn cháng shíjiān gēn wǒ zhēngdòu. ránhòu, wǒ zhǎngdà le,
‘They fought with me for a long time. Then, I grew up.’

Subject pronoun is a widely-studied sociolinguistic variant in variation studies (e.g. Bayley & Pease-Alvarez, 1996, 1997; Flores-Ferrán, 2002; Jia & Bayley, 2002; X. Li et al., 2012; Travis, 2007). The two major linguistic constraints are referent switch and subject person/number. Social factors include age, gender, occupation, and discourse context. Jia and Bayley (2002) found that Chinese subject pronoun use is influenced by discourse context. X. Li et al. (2012) found that older speakers/teachers/females tend to use overt pronouns in subject positions more than younger speakers/students/males. As to subject pronoun use by L2 Chinese learners, X. Li (2014) found it is influenced by various linguistic factors including referent switch, subject person/number/animacy, sentence type, referent specificity, discourse (conversation vs. elicited narrative), learner’s native language, gender, and proficiency level.

This study includes bigger datasets than previous studies and compares the general patterns of DE use and subject pronoun use by Chinese college students, Chinese teachers, advanced-level learners of Chinese, and the textbooks used by the learners. By doing so, we can gain a better understanding of what learners need to focus on in order to produce more sociolinguistically appropriate native-like language features, and how we as language instructors can help them.
4. Methodology

4.1 Data

The data in this study come from a larger project on the use of sociolinguistic variants by Chinese native speakers and CSL learners, and include four sources. The first source was a series of casual interviews between the author and 23 CSL learners (13 females and 10 males; mean age of 26) at two major universities in China. They were enrolled in high-intermediate or advanced Chinese classes, and came from four native language groups: English, Korean, Japanese, and Russian. The learners were interviewed twice in one semester with approximately ten weeks in between. The interviews covered casual topics such as favorite teacher, favorite book, experience in China, friends, and travelling experiences. Each interview lasted 45 minutes to an hour.

The second data source included the author’s conversations with 13 Chinese college students from one of the universities where the CSL learners were studying. Ten of these native speakers were from the northern Chinese province of Heilongjiang and three from other provinces of China, but all of them spoke fluent Mandarin, which is the focus of the study. This native speaker group consisted of six female and seven male students with an average age of 18.7. The native speakers were interviewed once for about 30 to 40 minutes. The interviews were carried out in Mandarin and the topics were similar to those of the CSL learner interviews.

The third data source was the L2 learners’ instructors in their Chinese classes. There were four instructors (three females and one male). Their classroom speech was audio-recorded. The instructors had been teaching Chinese formally for 34, 20, 4, and 4 years, respectively.

The fourth data source was four Chinese textbooks used in the learners’ classes. The reading texts in the lessons taught were used for analysis. Grammar and exercise sections were not examined. Most of these texts were selected from authentic newspaper articles, literary works, and magazines in China; therefore,

2. The four textbooks were the following:

the analysis of stylistic variation in these texts also manifests stylistic variation patterns in written Chinese.

In sum, this study collected and analyzed a total of 43 hours of the L2 learners’ conversational speech, eight hours of native speakers’ conversational speech, 16 hours of the instructors’ in-class speech, and 31 lesson texts from the textbooks.

4.2 Analytical approach

Aiming to examine the relationship between educational input (teacher speech and textbooks) and learner production of the target linguistic variants (particle DE and subject pronoun) as compared to the speech of native speaker peers, the researcher compared general patterns of variant use by learners, native speakers, teachers, and textbooks. Two aspects related to data analysis need to be clarified. First, previous studies that explored educational factors in the teaching of sociolinguistic variation examined how the variants were introduced and practiced in the textbooks, and how teachers used them (De Vito, 1991; Rehner & Mougeon, 2003). Since the participants in this study had high-intermediate and advanced levels of Chinese proficiency and had already been taught the target variants in their previous lower-level classes, this study analyzed how the variants were used by teachers and represented in the textbooks. Second, this study did not analyze variability within group (high-intermediate or advanced), because the study’s focus was to glean the general relationship between learner speech and educational input, with native speaker speech as the benchmark.

Linguistic variation analysis was conducted on the data using GoldVarb X (Sankoff, Tagliamonte, & Smith, 2005). This program provides frequency analysis indicating how often variants are used as well as the factor weights that indicate the likelihood of variant use. The application values set in the program were the applications of DE and subject pronoun. Therefore, a factor weight over .50 would indicate favoring the overt form (i.e., DE use and explicit subject pronoun) and below .50 would indicate favoring the null forms. An input value indicated the overall likelihood of variant use in overt form in the dataset analyzed, and a frequency value indicated the percentage of variant use in overt form. In addition, a one-way ANOVA was used to compare the frequencies between groups to see if there were any differencies, and if so, what the differences were.

4.3 Coding

As I mentioned earlier, this study was part of a larger sociolinguistic study that included a larger number of independent variables. For the analysis of the particle DE, the independent variables included various DE functions, gender, formality,
proficiency level, length of stay in China, native language, and proficiency level. For the analysis of the subject pronoun, the independent variables included subject person/number, subject-referent switch, sentence type, native language, length of stay in China, gender, discourse context, and proficiency level. All of the actual and potential occurrences of target language features (use and nonuse of DE; overt and null pronoun in subject positions) were identified, and coded with some exceptions. For DE, coding exceptions included the use of proper nouns, reformations or repetitions, false starts, lexicalized expressions, multiple modifiers, and error cases. For subject pronoun use, coding exceptions included false starts, reformations or repetitions, and obligatory and prohibited cases. The coding was done by one researcher and then checked by another researcher. The author examined all the discrepancies and discussed them with the second researcher to reach agreement. All three researchers were Chinese native speakers who were trained and experienced in coding Chinese natural speech data. The interrater reliability was .91 for DE coding and .92 for subject pronoun coding. Only the results that pertain to this study will be reported in the following.

5. Results

5.1 Use of the particle DE

Four separate runs of variation analyses for DE were conducted with GoldVarb X on the four datasets: native speakers’ speech, CSL learners’ speech, teachers’ speech, and textbook input. The results showed that although all four sources favor the overt form of DE generally, the overall likelihood of DE use in learner speech (.83), teacher speech (.84), and textbooks (.85) is much higher than DE use in native speaker speech (.63). A one-way ANOVA was conducted to compare the frequencies across all four groups, and the results indicated that there was a statistically significant difference between groups: $F(3, 24044) = 294.06 \ (p < .001)$. A Tukey post-hoc test revealed that frequency of overt DE use was significantly lower in the native speakers’ speech than in the CSL learners’ speech, teachers’ speech, or textbook input ($p < .001$). There were no statistically significant differences in frequency among the learners’ speech, teacher input, and textbook input. Table 1 presents the results.

Since DE has many functions, further examination of DE use in different function environments by different groups was conducted to see where the differences lay. Table 2 presents token frequency of explicit DE use by the four groups. The results show several trends. First, consistent with the results shown in Table 1 and Figure 1, learners used DE at a much higher rate than their native speaker peers,
especially in four DE functions: phrase as modifier, DE structure followed by a demonstrative or number + classifier, genitive marker, adjective or noun as modifier.

Second, learners’ patterns of DE use followed closely those of their teachers’ input and textbook input in all grammatical functions except for the verb as modifier (57%). Reexamination of the textbook data shows that 60% of the verb-as-modifier cases are four-syllable expressions (e.g., 读书 习惯, dūshū xíguàn, reading habits). Four-syllable expressions are unique in Chinese, especially in Chinese writing. The cases in this study are not lexicalized expressions, but the verb modifiers acquire a nominal feature and behave like noun modifiers (X. Li, 2010a).

Table 1. DE use in Chinese NSs’ speech, CSL learners’ speech, teacher input, and textbook input (general pattern)

<table>
<thead>
<tr>
<th></th>
<th>Chinese NSs’ speech</th>
<th>CSL learners’ speech</th>
<th>Teacher input</th>
<th>Textbook input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall frequency (%)</td>
<td>55</td>
<td>77</td>
<td>76</td>
<td>75</td>
</tr>
<tr>
<td>Input value</td>
<td>.63</td>
<td>.83</td>
<td>.84</td>
<td>.85</td>
</tr>
<tr>
<td>Number of tokens</td>
<td>4,505</td>
<td>14,220</td>
<td>2,458</td>
<td>2,535</td>
</tr>
</tbody>
</table>

Note: Data of teacher input and textbook input come from X. Li (2010a)

Table 2. DE use in Chinese NSs’ speech, CSL learners’ speech, teacher input, and textbook input (in different functions) (token frequency)*

<table>
<thead>
<tr>
<th></th>
<th>Chinese NSs’ speech</th>
<th>CSL learners’ speech</th>
<th>Teacher input</th>
<th>Textbook input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominalization¹</td>
<td>93</td>
<td>93</td>
<td>98</td>
<td>99</td>
</tr>
<tr>
<td>Relative clause²</td>
<td>91</td>
<td>99</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Verb³</td>
<td>85</td>
<td>92</td>
<td>91</td>
<td>57</td>
</tr>
<tr>
<td>Phrase⁴</td>
<td>76</td>
<td>90</td>
<td>83</td>
<td>93</td>
</tr>
<tr>
<td>Followed by DorNC⁵</td>
<td>42</td>
<td>78</td>
<td>63</td>
<td>72</td>
</tr>
<tr>
<td>Genitive⁶</td>
<td>33</td>
<td>84</td>
<td>72</td>
<td>73</td>
</tr>
<tr>
<td>Adj+Noun⁷</td>
<td>26</td>
<td>56</td>
<td>55</td>
<td>48</td>
</tr>
</tbody>
</table>

Notes 1. Nominalization marker
2. Relative clause as modifier
3. Verb as modifier
4. Phrase as modifier
5. Followed by a demonstrative or number + classifier
6. Genitive marker
7. Adjective or noun as modifier

* Data of teacher input and textbook input come from X. Li (2010a)
5.2 Subject pronoun

Four runs of variation analyses for subject pronoun use were conducted using GoldVarb X on the four datasets from native speakers’ speech, CSL learners’ speech, teacher input, and textbook input. The results showed that generally CSL learners (.60) and teachers (.60) favored overt subject pronoun use, but native speakers (.34) and textbooks (.26) favored null form. A one-way ANOVA was conducted to compare the frequencies across all four sources and the results revealed a statistically significant difference between groups, $F(3, 25170) = 518.31$ ($p < .001$). A Tukey post-hoc test revealed that the frequency of overt pronouns was significantly lower in native speakers’ speech (35.9%) than in learners’ speech (59.4%) or in teachers’ input (58.9%) ($p < .001$), but the frequency was higher than in textbook input (29.9%) ($p < .001$). Learners and teachers used overt subject pronouns at a significantly higher rate than textbooks ($p < .001$). There was no statistically significant difference in the frequencies of overt subject pronouns between learners’ speech and teacher input. Table 3 presents the frequency results and the input values.

Table 4 presents the results of a further examination of subject pronoun use by different groups in different subject-referent switch conditions. Again, the frequency values refer to the percentage of explicit subject pronoun use. The patterns generally corroborate the results shown in Table 3, which are as follows: (1) Teachers used the overt language forms more than native speakers did; (2) Learners demonstrated much more frequent use of overt language forms compared with
their native speaker peers; (3) Textbooks included overt subject pronouns least frequently; and (3) Patterns found in the learners’ speech followed those in the teacher input closely, but frequencies in the learners’ data were generally higher than those in the teacher input.

There are two conditions in the textbook input that go against the general pattern – “same underlying” (see Example 12) (63%) and “fronted topic” (see Example 15) (50%) – and are higher than native speaker frequencies (51% and 31%). Further examination of the data shows that these two conditions occur most of the time in the dialogues shown in the textbooks – 81.3% (13/16) for “same underlying” and 66.7% (4/6) for “fronted topic.”

Table 3. Subject pronoun use in Chinese NSs’ speech, CSL learners’ speech, teacher input, and textbook input (general pattern)

<table>
<thead>
<tr>
<th></th>
<th>Native speakers’ speech</th>
<th>CSL learners’ speech</th>
<th>Teacher input</th>
<th>Textbook input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall frequency (%)</td>
<td>35.9</td>
<td>59.4</td>
<td>58.2</td>
<td>29.9</td>
</tr>
<tr>
<td>Input value</td>
<td>.34</td>
<td>.60</td>
<td>.60</td>
<td>.26</td>
</tr>
<tr>
<td>Number of tokens</td>
<td>4,321</td>
<td>13,354</td>
<td>2,661</td>
<td>3,313</td>
</tr>
</tbody>
</table>

Note: Data of native speakers and CSL learners come from X. Li (2014)

Table 4. Subject pronoun use in Chinese NSs’ speech, CSL learners’ speech, teacher input, and textbook input (different referent switch conditions) (token frequency)

<table>
<thead>
<tr>
<th></th>
<th>Native speakers’ speech</th>
<th>CSL learners’ speech</th>
<th>Teacher input</th>
<th>Textbook input</th>
</tr>
</thead>
<tbody>
<tr>
<td>Switch</td>
<td>52</td>
<td>71</td>
<td>70</td>
<td>52</td>
</tr>
<tr>
<td>Partial overlap</td>
<td>58</td>
<td>78</td>
<td>73</td>
<td>41</td>
</tr>
<tr>
<td>Other – subject</td>
<td>45</td>
<td>72</td>
<td>74</td>
<td>33</td>
</tr>
<tr>
<td>Same underlying</td>
<td>51</td>
<td>52</td>
<td>78</td>
<td>63</td>
</tr>
<tr>
<td>Fronted topic</td>
<td>31</td>
<td>85</td>
<td>56</td>
<td>50</td>
</tr>
<tr>
<td>Object – subject</td>
<td>42</td>
<td>65</td>
<td>47</td>
<td>27</td>
</tr>
<tr>
<td>No switch</td>
<td>24</td>
<td>46</td>
<td>38</td>
<td>18</td>
</tr>
</tbody>
</table>

Note: Data of native speakers and CSL learners come from X. Li (2014)
6. Discussion

Three aspects of the findings in this study merit discussion. First, the teachers used the overt forms of the target sociolinguistic variants at a significantly higher rate than the native speakers did in their everyday conversations. Since the teachers’ speech used in this study was collected during their classroom teaching, it is possible that they accommodated the learners’ speech by adopting a clear, direct, and transparent speech style manifested by the use of more overt forms of language features (DE in optional contexts and overt subject pronoun). As outlined earlier, DE specifies the relationship between the modifier and the head noun or noun phrase. Subject pronoun, on the other hand, clearly indexes the referent in the context. Providing the overt forms of subject surely helps learners’ understanding of the referents. This finding is in line with previous studies that found a mismatch of sociolinguistic variant use in vernacular speech and teacher speech (X. Li, 2010a, 2014; Mougeon, Nadasdi, & Rehner, 2010; Mougeon, Rehner, & Nadasdi, 2004; Nadasdi, Mougeon, & Rehner, 2005; Rehner & Mougeon, 2003; Rehner, Mougeon, & Nadasdi, 2003). Corroborating these studies’ findings, this study found that the way native speakers use the language in colloquial speech is not authentically represented in teachers’ classroom speech.

The second major finding is that learners tend to overuse the overt forms of the target features (i.e. DE use and overt subject pronoun) compared with their native speaker peers. This pattern seems to align with teachers’ speech patterns. Previous studies also found that learners overuse overt forms in optional contexts (Duff &
Li, 2002; Polio, 1995). For example, Duff and Li (2002) found that native speakers tend to be more conservative in marking LE in optional contexts than learners. Polio (1995) found that learners tend to produce overt pronouns in oral Chinese much more frequently than native speakers. Learners might choose to produce overt forms of language structures in optional contexts for the sake of correctness and clarity. In addition, learners usually consider their teachers the language authority and might adhere to their speech patterns unconsciously. Previous studies found significant impact of teacher speech on learners’ use of sociolinguistic variants (X. Li, 2010a, 2014; Mougeon, Nadasdi, & Rehner, 2010; Mougeon, Rehner, & Nadasdi, 2004; Nadasdi, Mougeon, & Rehner, 2005; Rehner & Mougeon, 2003; Rehner, Mougeon, & Nadasdi, 2003) and the findings of this study support that.

Finally, unlike teacher input, another classroom input examined in this study – textbooks – demonstrated mixed results. The patterns of DE use in textbooks and teacher speech are very similar. However, the subject pronoun use in textbooks is significantly less frequent than that in teacher speech (except for two coreference conditions – “same underlying” and “fronted topic”) or in speech by native speakers. This finding is slightly different from previous studies, which found a correspondence between textbook input and learners’ use of sociolinguistic variants (Etienne & Sax, 2009; X. Li, 2010a; Rehner & Mougeon, 2003). This might be partially due to the difference in focus of this study’s textbook investigation. Previous studies examined lower-level language textbooks, analyzing their grammar, constructed dialogues, and practice sections (De Vito, 1991; Etienne & Sax, 2009). In contrast, participants in this study were higher-level L2 Chinese learners and hence more advanced level textbooks were analyzed. In addition, this study analyzed only the lesson texts appearing in the textbooks and did not analyze grammar and practice sections, because the target variants had already been introduced in the participants’ previous classes, and the practice sections in the textbooks did not focus on these two linguistic features. On another note, though, in the current study, the textbook texts examined for DE use and subject pronoun were the same but the patterns found were different. For DE use, learners’ patterns aligned with the patterns in textbooks, but for subject pronoun use, they were significantly different. The frequency of subject pronoun use in textbooks was much lower than in learner speech, and even lower than in native speaker speech. I speculate that this is due to a significant difference between Chinese spoken discourse and written discourse in the use of anaphoric reference. The majority of the texts in the textbooks analyzed in this study came from authentic Chinese magazines, newspapers, and literary works. Previous studies have revealed the use of anaphoric reference as a main difference between oral and written discourse (Beaman, 1982; Chafe, 1982; Christensen, 2000). Chafe (1982) and Beaman (1982) claimed that the spontaneous and interactive nature of spoken discourse is the primary drive
for its difference from written discourse. More planning time allows the more frequent appearance of topic chains in Chinese written discourse. In topic chains where multiple clauses occur under the same topic, subjects tend to be omitted (W. Li, 2004). Christensen (2000) confirmed this tendency by showing that the omission of pronouns or noun phrases is more prevalent in Chinese written discourse than in oral discourse. Less frequent appearance of subject pronouns in the textbooks analyzed in this study can be attributed to this feature of Chinese written discourse involving topic chains. It seems that, in this study, learner speech followed teacher speech pattern more than textbooks. The different patterns of variant use in textbooks found in this study are actually the patterns represented in authentic written Chinese, instead of the artificially-constructed dialogues typical in the lower-level textbooks analyzed in previous studies.

7. Conclusion and instructional implications

As shown in this study, learners’ use of sociolinguistic variants aligned with teacher speech patterns more than with those found in textbooks. The findings bear important implications for foreign language education.

First, corroborating existing literature (Cohen, 2008; Etienne & Sax, 2006; 2009; X. Li, 2010a, 2014; Rehner & Mougeon, 2003; Taguchi, 2011, 2015), sociolinguistic variants need to be incorporated into foreign language education and explicitly instructed. Leaving the task entirely to the learners to figure out the appropriate use of the variants by listening to their teachers, reading the textbooks, and interacting with the native speakers is not effective, and can sometimes be misleading. Once we determine that it is necessary to incorporate the sociolinguistic aspect of language learning into the curriculum despite its complexity, we can move on to explore what to teach and how to teach it.

Second, clear goals need to be set up regarding which language modality we are teaching, because the patterns of use of sociolinguistic variants are different in speech and writing, as shown in this study and previous literature. These goals are needed for both material design and teaching itself. Take subject pronoun use, for example: overt pronouns occur much less frequently in written Chinese; therefore, when teaching this language structure, one must carefully analyze authentic Chinese texts for the representation of this linguistic form.

Third, it is important to design appropriate materials for target language forms, in which native speakers demonstrate variability, and explicitly teach these language forms as well as their pragmatic, stylistic, and social meanings. Existing literature has explored different ways to teach sociopragmatic or sociolinguistic variation. Taguchi (2015) conducted a comprehensive review of instructed
pragmatics and found two general trends: (1) explicit instruction with metapragmatic information and production practice are effective for teaching pragmatics; and (2) implicit instruction with activities that involve learners’ noticing and processing of the target structures are more effective than mere exposure to input. Therefore, developing a high level of awareness is necessary in order to acquire complex pragmatic and sociolinguistic aspects of the target language. Lyster (1994) explored the use of a functional-analytic approach to teaching the French address forms *tu* and *vous*. Students were guided to explicitly learn and practice target variants. He found that this method significantly improved students’ ability to recognize the appropriate use of address forms and to use the variants appropriately in both speaking and writing. Cohen (2008) argued for the importance of self-learning strategy training: “having the teachers give initial guidance and then leaving the actual learning of pragmatics to the students” (p. 231). Lemmirich (2010) proposed a web-based awareness-raising approach to teach sociopragmatic variation in L2 German. These are intriguing studies that provide valuable information. However, there is no textbook that incorporates empirical findings on sociopragmatic/sociolinguistic variation in a systematic fashion. Systematic instructional materials need to be developed based on empirical findings about native speaker patterns of language use in different contexts, with consideration to learners’ levels and target language features. In order to do this, close collaboration among language instructors, researchers, and textbook writers is needed. In addition, textbook materials need to incorporate information about sociolinguistic/sociopragmatic features in the target language, including not only explicit explanations of the features but also presentations of authentic examples of the features in texts and exercises. In addition, different functions of target variants also need to be introduced at different stages of learning based on their levels of difficulty, since learners seem to acquire certain variants more easily than others because of varying features in different variant functions. For example, “same underlying” in subject pronoun use is acquired first, and “fronted topic” seems to be more difficult, as demonstrated by the frequency differences between learners and native speakers in this study (1% – same underlying; 54% – fronted topic). It seems that variant functions that involve more discourse management are more difficult to acquire and need more assistance.

Finally, the best way to teach sociolinguistic/sociopragmatic variation is appropriate inclusion of authentic materials into the curriculum. For example, Etienne and Sax (2006) advocated using films to teach stylistic variation. When choosing authentic materials, a variety of factors need to be examined, such as learners’ proficiency levels, material content, student interest, target language structures, and teaching goals. Careful design of activities based on students’ language levels and teaching goals is also important.
8. Future directions

Turning back to the influence of textbooks on learner use of variants, the findings of this study bring up some intriguing research questions. How much does variant use in reading influence learner speech? How large a role do learner attention and noticing play in the acquisition of the variants? Do we need to consider modality when trying to teach students sociolinguistic variants? Do we need to investigate the use of sociolinguistic variants in both speech and writing in the native language? All these questions not only direct our future research but also the design of instructional materials that teach sociopragmatic/sociolinguistic variants. We need systematic investigation in different modalities so that we know what patterns learners need to acquire in order to produce authentic language in different social situations.

References


本文探讨汉语作为第一语言和第二语言的总体风格变体规律。研究重点是形态句法虚词“的”和主语代词。数据来自于 13 个汉语母语使用者、四位汉语教师、23 个汉语学习者和四本汉语教材。变体性分析和频率分析结果显示四个主要规律。第一，教师课堂上的语言中使用风格变体的频率明显高于汉语母语使用者在对话中的使用频率。第二，和同龄母语使用者比较，汉语学习者有过度使用风格变体的趋势。第三，学习者风格变体使用规律和他们老师的规律相近。最后，和第三点不同的是，教材中风格变体的使用规律和学习者的规律比较，结果不一。“的”使用，学习者和教材中的规律显著相近，但是主语代词使用规律则不然。然后作者讨论了本研究对于汉语口语和笔语中风格变体规律研究的含义，还有对汉语外语教学的启示。

**关键词:** 风格变体, 社会语言学, 汉语外语教学

**Author’s address**

Xiaoshi Li  
Department of Linguistics and Germanic, Slavic, Asian and African Languages  
Michigan State University  
B259 Wells Hall, 619 Red Cedar Rd.  
East Lansing, MI 48824  
xli@msu.edu