Abstract

To understand how daily usages can shape the gradual changes of both hěn, a prototypical intensifier in Mandarin Chinese, and the construction hěn X, the study aims to investigate the syntactic and semantic behaviors of hěn X constructions in spoken corpora. The conversational data from the NCCU Corpus of Spoken Chinese and a Taiwan Public Television show Bring Up Parents are extracted and analyzed, focusing in particular on the syntactic categories of X, the grammaticalization of hěn, and the lexicalization of hěn X. Several findings are found. First, the syntactic and semantic distributions of the data from both corpora are quite consistent. While adjectives and stative verbs still claim the majority of X, new categories are discovered, showing host expansion of X. In addition to words, phrases and clauses can play the role of X. The increase of the flexibility and complexity of X demonstrates the gradual grammaticalization of hěn. Moreover, some instances of hěn X can be used as a unit to modify other grammatical constituents, showing a certain degree of lexicalization. When hěn X is fused as a unit, hěn is obligatory, not only indicating a degree but also highlighting the characteristics of X. The analysis shows that the nature of spoken materials enhances the subjectivity of hěn X. The findings of hěn X in spoken corpora can be applied to linguistic studies and Mandarin teaching.

1 Background

Hěn X constructions, often employed in both spontaneous speeches and written texts, have undergone syntactic and semantic changes. In addition to modifying common adjectives and verbs, the degree adverb hěn collocates with various types of words. While many studies have discussed the history of hěn, the development of its degree-specifying function and the expansion of X, in general, the main claim is that hěn, as a prototypical intensifier in Mandarin Chinese, has grammaticalized into a grammatical marker in conjunction with its gradual loss of lexical meaning but its gaining of subjective evaluation (Chui, 2000; Lin, 2009; Tseng, 2010; Bai and Zhao, 2007, Chen and Tsai, 2008, Liu and Chang, 2012). Among previous studies, few have discussed hěn X constructions in spoken data although they are used more and more frequently in daily conversations with hěn indicating a higher degree than normal states and with X expanding to various syntactic categories. The usages of hěn X constructions in spoken corpora deserve further exploration. To understand how daily usages can shape the gradual changes of hěn X, the study aims to investigate the syntactic and semantic behaviors of hěn X constructions in two different spoken corpora, focusing in particular on examining the syntactic categories of X, the grammaticalization of hěn, and the lexicalization of hěn X.
2 Grammaticalization and Lexicalization

Brinton and Traugott (2005:96-99) emphasize the highly interactive relation between grammaticalization and lexicalization in language change. Lexicalization refers to a word formation process in which a new lexical item is produced with its structural and semantic properties not completely derivable from the components of the word formation pattern. The output of such a process forms a gradient continuum of complexity, ranging from fixed or idiomatic phrases (L1), to compounds and derived forms (L2), and to lexical simplexes and idiosyncratic fossilized forms (L3). The degree of lexicalization within a word increases along with the loss of its grammatical and semantic element features, and lexicalization processes form a gradient continuum by the three levels of lexicality L1, L2, and L3. Grammaticalization, on the other hand, refers to a process whereby lexical items or constructions are used to serve a grammatical function in certain linguistic contexts, and become more grammatical by obtaining more grammatical functions and expanding their host-classes. Grammaticalization processes also form a gradient continuum on a scale of grammaticality G1, G2, and G3. Brinton and Traugott (2005) point out the differences and similarities of the two processes. Lexicalization integrates existing forms to serve as members of a major category, but grammaticalization involves decategorization of forms from major categories to minor ones to serve grammatical functions. However, both processes involve a decrease in syntactic or semantic compositionality and an increase in fusion. The analysis of this study indicates that both processes are involved in hěn constructions as will be shown below.

3 Methodology

3.1 Data

The data are taken from The NCCU Corpus of Spoken Chinese (Chui and Lai, 2008) and the TV interview show Bring Up Parents (爸媽囧很大) from Taiwan Public Television Service Foundation. The NCCU Corpus of Spoken Chinese, an online open access spoken data, consists of around 9 hours of 27 Mandarin daily conversations with two or three Mandarin-speaking adults. Bring Up Parents is a TV program containing interviews and conversations of both parents and their sons or daughters. The episodes from July to December in 2013 were selected. The First and Eighth episodes in every month were extracted, totaling 12 episodes of 12 hours. In total, 805 tokens of the NCCU corpus, and 870 tokens of the TV show will be examined.

3.2 Data coding

The tokens of hěn X are coded regarding the syntactic categories of X, the number of words of X, the grammatical function of hěn X, and the meaning of hěn X. The procedure is shown below.

(A) Syntactic category: Analyzing the syntax category of X as NOUN, VERB, ADJECTIVE, ADVERB, PRONOUN, or PREPOSITION.

(B) Word number: Counting word number of X following hěn.

(C) Grammatical function: Indicating the grammatical functions of hěn X as SUBJECT, PREDICATE, OBJECT, ATTRIBUTIVE, ADVERBIAL, or COMPLEMENT.

(D) Semantic function: Observing the contexts of hěn X, and analyzing the meaning.

4 Results and Discussion

4.1 Results

The result is shown in Appendix 1. The distributions are quite similar in the two spoken corpora. Regarding syntactic categories, adjectives and stative verbs claim the majority of X. However, new categories such as nouns, relation verbs, modal verbs, adverbs, prepositions, and pronouns are found, showing host expansion of X. Regarding the syntactic functions of hěn X, serving as a predicate displays a major part in the distribution, and all complements are modal complements. Serving as subjects or objects are rare. The distributions of attributives and adverbials are alike. The only difference is that restrictive attributives are used more frequently in TV shows, and mostly is the lexicalized form hěnduō (很多 hen-many; ‘frequent’). Regarding word numbers, X is found to contain one or two words in majority. When X contains three words, the string is usually a relational verb. The syntactic behaviors and meanings of hěn X vary in actual usages. When
used to modify adjectives and stative verbs, hěn objectively indicates the degree of shapes or quality. Consider the examples from (1) to (3). Long or thin in shape, hard or soft in property, and sour, bitter or stinky in senses may correspond to temper, cultural, life and emotions. Notice that an interesting feature of hěn is that sometimes it seems to be semantically bleached, becoming an obligatory marker. In these three examples, the three predicates cannot stand alone without hěn; however, there is no intensification present in the sentences. The construction of hěn X is lexicalized to some extent with metaphorical meaning extended from the whole construction.

(1) 有時候老公會問的 很細，然後我就會覺得 很煩
Yǒushíhòu lǎogōng huì wèn de hěnxì ránhòu wǒ jiù huì juéde hěnfán (sometimes-husband-will-ask-COMP-very detailed-then-I-will-feel-very annoyed)
‘When my husband asks for too many details, I will feel very annoyed.’

(2) 生物老師 很硬，考的生物非常難。
Shēngwù lǎoshī hěnyìng kǎo de shēngwù fēicháng nán (biology-teacher-very tough-test-biology-very difficult)
‘Biology teachers are very tough, often giving students difficult tests.’

(3) 結婚 很苦，碰到很多波折。
jiéhūn hěnkǔ pèngdào hěnduō bōzhé (marriage-very bitter-bump into -many-frustrations)
‘Marriage is bitter; I bump into many frustrations.’

Due to its property-modifying function, hěn can collocate with X denoting state or property. Thus, the host X can further extend to nouns, action verbs, relational verbs, modal verbs, pronouns, adverbs, and prepositions, which could not be modified before. The meaning is metaphorical with hěn X lexicalized as a fused unit. When modifying a noun, hěn will trigger the appropriate semantic property contingent to linguistic contexts. For instance, in the case of hěntǔ (很土 hen-earth; ‘out of fashion’), the concrete property of earth changes into projecting the purpose and function of the substance. Also, a metaphorical meaning is observed in the compound. Hěn highlights the fixed and invariable property of earth to metaphorically express a pejorative extended meaning--out of fashion. Cases that carry similar metaphorical meanings are hěnbāgǔ (很八股 hen-stereotyped; ‘hackneyed’) or hěnkǒuhào (很口號 hen-slogan; ‘like a slogan’).

Grammaticalization of hěn occurs, with hěn extending to modify a noun denoting abstract orientation, as in the following example:

(4) 天母...房子都打 很下面，然後下面都是停車場...
Tiānmǔ fāngzi dōu dǎ hěnxiàmiàn ránhòu xiàmiàn dōu shì tíngchēchǎng (Tianmu-house-all-build-very low-then-down-all-is-garage)
‘Houses in Tianmu are built farther under the ground for parking garages....’

The chunk hěnxiàmiàn denotes farther down the ground, and hěn emphasizes not only the degree of orientation but also the speaker’s subjective evaluation of the situation.

The categories of X can also expand to relational verbs, modal verbs and action verbs. One interesting example has to do with the co-occurrence of hěn with you X construction to emphasize a high degree above the average. For example, in (5), hěnyǒutónggǎn (很有同感 hen-have-same feeling ‘feel the same way’) is to emphasize the speaker’s feeling and thinking. And in (6) hěnyǒugǎnqíng (很有感情 hen-have-feelings; ‘have deep feelings’) is a grammatical unit to express the speaker’s emotions. Thus, hěn is employed to denote a high degree associated with you constructions.

(5) 我很有同感，我很贊同他說的。
wǒ hěnyǒutónggǎn wǒ hěn zàntóng tā shuō de (I-hen-have-the same feeling-I-hen-agree-what he has said)
‘I feel exactly the same way. I totally agree with what he has said.’

(6) 他沒有對我 很有感情。
tā méiyǒu duì wǒ hěnyǒugǎnqíng (he-not-dui-me-hen-have-feeling)
‘He didn’t have deep feelings toward me.’
When modal verbs indicating obligation collocate with *hěn*, the construction denotes speakers’ subjectivity toward judging facts and emotions, and the collocation with *hěn* strengthens the speakers’ subjectivity. The case *hěn bù yīnggāi* (*很不應該*; ‘really shouldn’t’) can illustrate. The case *hěn huì zhǔ* (*很會煮*; ‘really good at coking’) shows the speaker’s evaluation of someone’s talent in cooking whereas *hěn huì tānliànài de* (*很會談戀愛*; ‘good at handling romantic relationships’) carries the speaker’s evaluation of someone being good at romantic relationships. Furthermore, *hěn* can modify lexicalized action verb phrases, as in *hěn chīlì* (*很吃力*; ‘very laborious’), *hěn xiàrén* (*很嚇人*; ‘very scary’), or *hěn jiànyì* (*很建議*; ‘highly recommend’) and *hěn shuǎshuài* (*很耍帥*; ‘look very cool’). Notice that in these cases *X* is getting more and more lexicalized, and that *hěn* is obligatory, revealing its grammatical function from grammaticalization.

The host classes of *X* keep expanding to pronouns as *hěn* further grammaticalizes. For example, in (7) and (8), speakers express euphemism by employing *hěnnàge* (*很那個*; ‘really-you know -bad’). The two examples show that *nàge*, a deictic expression, indicates abstract events and states. The usage of *hěnnàge* euphemistically expresses speakers’ negative thoughts toward the states. The co-occurrence of *hěn* and pronouns carries strong subjectivity since what the deictic pronoun *nàge* refers to can only be understood by contexts.

(7) 自殺真的很那個…
*Zǐshā zhěndé hěnnàge* (commit suicide-really-then CL)
‘It is really bad to commit suicide.’

(8) 我覺得考試去看電影很那個。
*Wǒ juéde kǎoshì qù kàn diànyǐng hěnnàge* (I-feel-exam-go-see movies -hen-then CL)
‘I think it’s not very good to see movies before the exam.’

Finally, *X* can even expand to include prepositions. Prepositions are function words indicating relations. In examples (9) and (10), *hěn* is a grammatical marker, strengthening the degree of the head specified by the predicate; it serves to express the relation between the speaker and the role modified by the preposition.

(9) 很替你感到難過
*Hěn tì nǐ gǎn dào nánguò* (hen-for-you-feel-sorry)
‘feeling really sorry for you’

(10) 很向鼻子靠近
*Hěn xiàng bízi kàojìn* (hen-toward-nose-approach)
‘approaching really toward the nose’

### 4.2 Grammaticalization and Lexicalization of *hěn X*

The discussion has shown that *hěn* is a degree adverb, intensifying the degree of its head specified by *X*. Due to the nature of spoken materials, *hěn* is further grammaticized with *X* being further expanded to include longer strings of words and more complex syntactic structures. While serving various grammatical functions like predicates, attributives, adverbials, or complements, some instances of *hěn X* such as *hěnhǎo* (*很好*; ‘very good’), *hěnnán* (*很難*; ‘hard to…’), *hěnduō* (*很多*; ‘frequent’), and *hěnshǎo* (*很少*; ‘little’) are lexicalized as fused units ready to modify other constituents.

Interestingly, the process of lexicalization continues as in cases such as *hěnhào X’, hěnnán X’, hěnduō X’, hěnyǒu X’, and hěnxìang X’. Such findings indicate that grammaticalization and lexicalization are highly interrelated processes. Owing to the frequent use of *hěn* with *hǎo* ‘good’, *nán* ‘difficult’, and *duō* ‘many’, and *shǎo* ‘few’, their word boundaries gradually diminish, producing a new semantic unit. For example, *hěnnán* can indicate either evaluation of possibility or the degree of difficulty as in *hěnnánshuō* (*很難說*; ‘hard to say’) and *hěnnán yǒukòng* (*很難有空*; ‘hard to be available’). While *duō* and *shǎo* represent amounts, *hěnduō* or *hěnshǎo* indicate frequency when qualifying abstracts or states as in *bāng hěnduō māng* (*幫很多忙*; ‘often help a lot’) and *hěnshǎo tán zhèjiànshì* (*很少談這件事*; ‘seldom discuss this...’).
The meaning of ³nhao has shifted from evaluating good quality to assess possibility ³nhaozhui (很好追 hen-good-chase ‘easy to hit on’).

The empirical findings of this study support the integration of grammaticalization and lexicalization proposed by Brinton and Traugott (2005). These two processes, motivated by speakers’ needs in interaction, undergo instantaneous changes and reanalysis. Language forms are repeatedly conducted by interlocutors, and gradually become fossilized. These gradual changes are dynamic with indeterminacy, revealing speakers’ subjective attitudes in daily usages. The subjectivity of the development of ³n X is justified as ³n X constructions mostly serve grammatical functions as predicates, descriptive attributives adverbials, or modal complements. The following two figures summarize the analysis of ³n X constructions. Figure 1 shows the gradual expansion of X from more prototypical categories like adjectives and verbs to less prototypical ones like nouns, pronouns, and prepositions.

Figure 1. The expansion of X

Figure 2 indicates that while the host X is expanded, ³n is becoming more and more grammaticalized into a grammatical marker. Some ³n X constructions such as ³nhao X’, ³nnán X’, ³nduo X’, and ³nsao X’, have lexicalized into a unit due to its frequent usage in spoken data. These cases have also developed their evaluative and subjective meanings in the contexts.

5 Conclusion

Grammatical and semantic changes happen due to speakers’ needs. This current work inspects the structural and semantic changes of ³n as an intensifier, as well as the syntactic and semantic behaviors of ³n X constructions in spoken corpora through quantitative and qualitative methods. The conversational data from the NCCU Corpus of Spoken Chinese and a Taiwan Public Television show Bring Up Parents are extracted and analyzed. Several findings are found. First, the syntactic and semantic distributions of the data from both corpora are quite consistent. Due to the nature of spoken materials, X reveals host expansion, expanding to new categories including nouns, adverbs, prepositions and pronouns other than the prototypical adjectives and stative verbs. It can also include not only words but also phrases and clauses. The increase of the flexibility and complexity of X demonstrates further grammaticalization of ³n. When X keeps on expanding to other syntactic categories, ³n X is developing toward the direction of grammaticality with an increase of its productivity (Briton and...
Traugott, 2005). However, some instances of hěn $X$ become lexicalized units serving to modify other grammatical constituents. When hěn $X$ is fused as a unit, hěn becomes an obligatory grammatical marker, expressing a higher degree than normal and at the same time highlighting the features denoted by its host. And such fused constructions are developing toward the direction of lexicality with a decrease of productivity (Briton and Traugott, 2005). The dynamic and interactive nature of conversations enhances the subjectivity of hěn $X$, in contingent with the integration of the processes of grammaticalization and lexicalization. The findings of hěn $X$ in spoken corpora can be applied to linguistic studies and Mandarin teaching.

**References**


Pai, Hsiao-hung, and Chao, Wei. 2007. 漢語虛詞十五講, Han-yü hsü-tz'u 15 chiang [Function words in Chinese]. Peking University Publisher, Beijing.

### Appendix: Syntactic Category and Grammatical Function Distributions in the Corpora

<table>
<thead>
<tr>
<th>Syntactic category</th>
<th>Noun</th>
<th>Verb</th>
<th>Adjective</th>
<th>Pronoun</th>
<th>Adverb</th>
<th>Preposition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical function</td>
<td>Corpus Show</td>
<td>Corpus Show</td>
<td>Corpus Show</td>
<td>Corpus Show</td>
<td>Corpus Show</td>
<td>Corpus Show</td>
</tr>
<tr>
<td>Subject</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Predicate</td>
<td>9</td>
<td>7</td>
<td>112</td>
<td>156</td>
<td>443</td>
<td>329</td>
</tr>
<tr>
<td>Object</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Attributive</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>105</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Descriptive</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>56</td>
<td>87</td>
</tr>
<tr>
<td>Adverbial</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Restrictive</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3.35%</td>
<td>12.06%</td>
</tr>
<tr>
<td>Descriptive</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>9</td>
<td>31</td>
<td>65</td>
</tr>
<tr>
<td>Resultative</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Directional</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Potential</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Modal</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>82</td>
<td>82</td>
</tr>
<tr>
<td>Quantity</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>10</td>
<td>124</td>
<td>174</td>
<td>647</td>
<td>674</td>
</tr>
<tr>
<td>Percentage</td>
<td>1.49%</td>
<td>1.15%</td>
<td>15.40%</td>
<td>20.00%</td>
<td>80.37%</td>
<td>77.47%</td>
</tr>
</tbody>
</table>