The Syntactic and Semantic Analysis of *Hěn* X Constructions in Spoken Corpora

Lai, Huei-ling  
National Chengchi University  
Taipei, Taiwan  
hllai@nccu.edu.tw

Chen, Yen-ju  
National Taiwan Normal University  
Taipei, Taiwan  
born95101078@gmail.com

Hsu, Shao-chun  
National Chengchi University  
Taipei, Taiwan  
shaochun@nccu.edu.tw

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 hěn-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
    Yǒushíhòu lǎogōng huì wèn de hěnxì
    ránhòu wǒ jiù huì juéde hěnfán
    ‘When my husband asks for too many details, I
    will feel very annoyed.’

(2)  植物老師很硬，考的生物非常難。
    Shēngwù lǎoshī hěnyìng
    kǎo de shēngwù fēicháng nán
    ‘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 is bitter; I bump into many
    frustrations.’

(4) 天母...房子都打很下面，然後下面都是停車場...
    Tiānmǔ fāngzi dōu dǎ hěnxìmiàn
    ránhòu xiàmiàn dōu shì tíngchēchǎng
    ‘Houses in Tianmu are built farther under the
    ground for parking garages...’

The chunk hěnxìmiàn 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-
ocurrence of hěn with you X construction to
emphasize a high degree above the average. For
example, in (5), hěnyǒutónggǎn (很有同感
have-same feeling ‘feel the same way’) is to
emphasize the speaker’s feeling and thinking. And
in (6) hěnyōugǎnqíng (很有感情
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 hěnyōugǎnqíng (he-not-due-
    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 yǐnggāi* (很不應該 hen-not-should; ‘really shouldn’t’) can illustrate. The case *hěn huì zhǔ* (很會煮 hen-able-cook; ‘really good at coking’) shows the speaker’s evaluation of someone’s talent in cook whereas *hěn huì tánliànài* de (很會談戀愛 hen-able-romance; ‘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 hen-eat-strength; ‘very laborious’), *hěn xiàrén* (很嚇人 hen-scared; ‘very scary’), or *hěn jiànyì* (很建議 hen-recommend; ‘highly recommend’) and *hěn shuāshuài* (很耍帥 hen-able- handsome; ‘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 *hěnnáge* (很那個 hen-that-CL; ‘really-you know -bad’). These two examples show that *náge*, a deictic expression, indicates abstract events and states. The usage of *hěnnáge* euphemistically expresses speakers’ negative thoughts toward the states. The co-occurrence of *hěn* and pronouns carries strong subjectivity since what the deictic pronoun *náge* refers to can only be understood by contexts.

(7) 自殺真的很那個

*Zìshā zhèndé hěnnáge* (commit suicide-really-hen-that 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 hěnnáge* (I-feel-exam-go-see movies -hen-that 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) 你替我感到難過

*Hěn tì wǒ gǎn-dào nánɡuò* (hen-for-you-feel-sorry)
‘feeling really sorry for you’

(10) 向鼻子靠近

*Hěn xiànɡ bìzi 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 grammaticalized 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 *hěnhào* (很好 hen-good; ‘very good’), *hěnnán* (很难 hen-hard; ‘hard to…’), *hěnduō* (很多 hen-many ‘frequent’), and *hěnshāo* (很少 hen-little ‘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 *hěnxiatang X*. Such findings indicate that grammaticalization and lexicalization are highly interrelated processes. Owing to the frequent use of *hěn* with *hào* ‘good’, *nán* ‘difficult’, and *duō* ‘many’, and *shāo* ‘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 hen-hard-say ‘hard to say’) and *hěnnán yǒukòng* (很難有空 hen-hard-available ‘hard to be available’). While *duō* and *shāo* represent amounts, *hěnduō* or *hěnshāo* indicate frequency when qualifying abstracts or states as in *bāng hěnduō máng* (幫很多忙 ban-hen-many-help ‘often help a lot’) and *hěnshāo tān zhèjiànshì* (很少談這件事 hen-little-discuss-this-matter ‘seldom discuss this
The meaning of hěnhǎo has shifted from evaluating good quality to assess possibility hěnhǎozhuī (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 hěn X is justified as hěn X constructions mostly serve grammatical functions as predicates, descriptive attributives adverbials, or modal complements. The following two figures summarize the analysis of hěn 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. Synchronic clines of lexicality and grammaticality

Figure 2 indicates that while the host X is expanded, hěn is becoming more and more grammaticalized into a grammatical marker. Some hěn X constructions such as hěnhǎo X’, hěnnán X’, hěnduō X’, and hěnshǎo 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 hěn as an intensifier, as well as the syntactic and semantic behaviors of hěn 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 hěn. When X keeps on expanding to other syntactic categories, hěn 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**

Brinton, Laurel and Traugott, Elizabeth Closs. 2005. *Lexicalization and Language Change*. Cambridge University Press, Cambridge, UK.

Chen, Ying and Tsai, Zheng. 2008. *Adverbs Hěn and Zhēn*. 現代漢語虛詞研究與對外漢語教學, *Xian dai han yu xu ci yan jiu yu dui wai han yu jiao xue*, 2, 14-27.

Chui, Kawai. 2000. Morphologization of the Degree Adverb HEN. *Language and Linguistics*, 1(1):45-59.

Chui, Kawai, and Lai, Huei-ling. 2008. The NCCU Spoken Corpora: Mandarin, Hakka, and Southern Min. *Taiwan Journal of Linguistics*, 6(2): 119-144.

Hong, Jia-Fei and Huang, Chu-Ren. 2006. Using Chinese Gigaword Corpus and Chinese Word Sketch in Linguistic Research. In *Proceedings of the 20th Pacific Asian Conference on Language, Information and Computation (PALIC’20)*, Wuhan, China.

Lin, Wen-yin. 2009. The use of *Hěn* in Modern Chinese. In International Workshop on Theory and Practice in Mandarin Chinese Instruction 2009, Taipei, Taiwan.

Liu, Mei-Chun, and Chang, Chun. 2012. The Degree-evaluative Construction: Grammaticalization in Constructionalization. In Xing Zhiqun (ed.) *Newest Trends in the Study of Grammaticalization and Lexicalization in Chinese*, 115-148. Mouton de Gruyter, Berlin.

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.

Tseng, Wei. 2010. Degree highlighting in spontaneous speech. *Journal of Shaoxing University*, 30(3), 67-72.
### Appendix: Syntactic Category and Grammatical Function Distributions in the Corpora

| Grammatical function | Source | Noun | Verb | Adjective | Pronoun | Adverb | Preposition |
|----------------------|--------|------|------|-----------|---------|--------|-------------|
| Corpus               | Show   | Show | Show | Show      | Show    | Show   | Show        |
| Subject              |        |      |      |           |         |        |             |
| 0                    | 0      | 0    | 0    | 1         | 6       | 0      | 0           |
| 0.00%                | 0.00%  | 0.00%| 0.11%| 0.74%     | 0.00%   | 0.00%  | 0.00%       |
| Predicate            |        |      |      |           |         |        |             |
| 9                    | 7      | 112  | 156  | 443       | 329     | 4      | 1           |
| 1.11%                | 0.80%  | 13.91%| 17.93%| 55.03%    | 37.81%  | 0.50%  | 0.12%       |
| Object               |        |      |      |           |         |        |             |
| 0                    | 0      | 3    | 0    | 1         | 4       | 0      | 0           |
| 0.00%                | 0.00%  | 0.37%| 0.12%| 0.45%     | 0.00%   | 0.00%  | 0.00%       |
| Attributive          |        |      |      |           |         |        |             |
| Restrictive          |        |      |      |           |         |        |             |
| 0                    | 0      | 0    | 0    | 27        | 105     | 0      | 0           |
| 0.00%                | 0.00%  | 0.00%| 0.00%| 3.35%     | 12.06%  | 0.00%  | 0.00%       |
| Descriptive          |        |      |      |           |         |        |             |
| 1                    | 2      | 5    | 4    | 56        | 87      | 0      | 0           |
| 0.12%                | 0.22%  | 0.62%| 0.45%| 6.95%     | 10.00%  | 0.00%  | 0.00%       |
| Adverbial            |        |      |      |           |         |        |             |
| Restrictive          |        |      |      |           |         |        |             |
| 0                    | 0      | 1    | 3    | 1         | 2       | 0      | 0           |
| 0.00%                | 0.00%  | 0.12%| 0.34%| 0.12%     | 0.22%   | 0.00%  | 0.00%       |
| Descriptive          |        |      |      |           |         |        |             |
| 1                    | 1      | 2    | 9    | 31        | 65      | 0      | 16          |
| 0.12%                | 0.11%  | 0.24%| 1.03%| 3.85%     | 7.47%   | 0.00%  | 1.98%       |
| Compliment           |        |      |      |           |         |        |             |
| Resultative          |        |      |      |           |         |        |             |
| 0                    | 0      | 0    | 0    | 0         | 0       | 0      | 0           |
| 0.00%                | 0.00%  | 0.00%| 0.00%| 0.00%     | 0.00%   | 0.00%  | 0.00%       |
| Directional          |        |      |      |           |         |        |             |
| 0                    | 0      | 0    | 0    | 0         | 0       | 0      | 0           |
| 0.00%                | 0.00%  | 0.00%| 0.00%| 0.00%     | 0.00%   | 0.00%  | 0.00%       |
| Potential            |        |      |      |           |         |        |             |
| 0                    | 0      | 0    | 0    | 0         | 0       | 0      | 0           |
| 0.00%                | 0.00%  | 0.00%| 0.00%| 0.00%     | 0.00%   | 0.00%  | 0.00%       |
| Modal                |        |      |      |           |         |        |             |
| 1                    | 0      | 1    | 1    | 82        | 82      | 0      | 0           |
| 0.12%                | 0.00%  | 0.12%| 0.11%| 10.18%    | 9.42%   | 0.00%  | 0.00%       |
| Quantity             |        |      |      |           |         |        |             |
| 0                    | 0      | 0    | 0    | 0         | 0       | 0      | 0           |
| 0.00%                | 0.00%  | 0.00%| 0.00%| 0.00%     | 0.00%   | 0.00%  | 0.00%       |
| Total                | 12      | 10   | 124  | 174       | 647     | 674    | 4           |
| Percentage           | 1.49%   | 1.15%| 15.40%| 20.00%    | 80.37%  | 77.47% | 0.50%       |
|                      | 2.11%   | 0.80%| 0.12%| 0.45%     |         |        |             |