

## Original Paper

# Functional Analysis of the “the X is” Construction: A Grammaticalization Perspective

Haixia Wang<sup>1</sup>

<sup>1</sup> School of English Studies, Beijing International Studies University, China

Received: October 03, 2025    Accepted: November 22, 2025    Online Published: December 10, 2025  
doi:10.22158/eltls.v7n6p142    URL: <http://dx.doi.org/10.22158/eltls.v7n6p142>

### Abstract

*The different propositional meanings and degrees of grammaticalization of head nouns (X) in the source structure are the root cause of the marked differences in the synchronic usage of the “the X is”-type expressions. The grammaticalization gradience of typical expressions conforming to this structure may be presented from high to low as follows: the thing is > the truth is/the fact is > the trouble is/the point is/the problem is > the reason is/.... Among them, the thing is is grammaticalized to a relatively high degree, and thus is more frequently used as a discourse marker; the trouble is, the point is and the problem is are grammaticalized to a relatively lower degree, and therefore are primarily used as comment clauses; most expressions of this type as in the case of the reason is, are minimally grammaticalized and retain their propositional meaning as matrix clauses.*

### Key words

*“the X is” construction, source structure, matrix clause, discourse marker, gradience of grammaticalization*

## 1. Introduction

The relevant research on discourse markers is mostly focused on functional investigations of individual cases, with less attention paid to discourse markers that appear in series. Moreover, the few previous studies on series-occurring discourse markers have largely been on the items composed of first- and second-person pronouns and verbs (Thompson & Mulac, 1991; Brinton, 2008; Cao, 2010). The “the X is” type of discourse markers, such as *the thing is*, *the fact is*, and *the truth is*, belongs to one of the few series of objects originating from third-person constructions. Although the terminology used varies, previous studies have largely noted the distinct discourse marker features and functions of such expressions (Crystal & Davy, 1975; Biber, 1999; Schmid, 2000). However, functional examination alone is insufficient to reveal the essential features, formation mechanisms, and interrelationships

among members of this type of discourse markers. A more comprehensive account of this issue is by Aijmer (2007), who proposes that fixed or semi-fixed units such as *the fact is that* and *the truth is that* can be defined as collocational frameworks from the perspective of emergent grammar, allowing specific local transformations, deletions, additions. The functions of different expressions are context-dependent. *The fact is*, *the truth is*, *the thing is*, and other expressions that contain epistemic head nouns (rather than all expressions that conform to this structure) reflect features of grammaticalization. This study explores the following issues based on Aijmer's research: what is the relationship between the various forms of "the X is" (e.g., *the thing is that*, *the thing is*, *thing is*, etc.) and their functions for a given head noun "X"; for different head nouns, what are the differences and connections between the "the X is" type of expressions, and how should they be interpreted, so as to provide a new perspective for understanding the "the X is" series of discourse markers.

## 2. Research Methods and Objects

Discourse markers can be seen as the result of grammaticalization (Traugott, 1995; Brinton, 2008; Prévost, 2011; Brems & Davidse, 2010). This study adopts Brinton's "matrix clause hypothesis" (2008) to analyze the evolution of the "the X is"-type expressions and, based on this framework, describes the synchronic differences in their usage.

A natural way to determine the grammaticalization stage of "the X is" is to use diachronic data to examine its evolution process. Considering the difficulties in obtaining diachronic data, this article proposes a combination of synchronic and diachronic approaches. The different uses of the items concerned in contemporary English (i.e., their marker use and non-marker use) can be taken as the chief stages involved in the process of grammaticalization. The underlying assumption is that the diachronic development is synchronically manifested in its different uses. In this way, the grammaticalization process of "the X is" can be possibly restructured so as to determine the relationship between its form and meaning.

Brinton's hypothesis summarizes the evolution path of the first- and second-person matrix-clause-type items as "matrix [subject + verb (present tense)] > indeterminate: matrix or parenthetical > comment clause", in which comment clauses are defined as discourse markers with a clausal structure that differ from other discourse markers by their clausal origin. Unlike Brinton, however, this study regards comment clauses and discourse markers as different phases because they are significantly different in terms of scope and function (Wang, 2014). While the former conveys the speaker's comment on the content or style of a sentence or clause, the latter works on the textual and (or) interpersonal level in discourse. As a result, the evolution process of "the X is" may be described as "matrix clause [subject + verb (present tense)] > indeterminate: matrix or parenthetical > comment clause > discourse markers". Based on the process of grammaticalization, Table 1 gives the possible grammaticalization stages and the corresponding functions and forms of the "the X is" type of expressions. The degree of grammaticalization of "the X is" may vary from one head noun (X) to another, and its usage and form

may vary accordingly. As to the more grammaticalized “the X<sub>4</sub> is” and “the X<sub>5</sub> is”, the importance of their various functions they serve may be different.

**Table 1. Grammaticalization Process of the “the X is” Construction**

Function	Matrix clause	→	Matrix clause comment clause	or	→	Comment clause	→	Discourse marker
Form	the X is+that-clause	→	the X is+that-less clause		→	the X is, that-less clause	→	(the) X is, that-less clause
X <sub>1</sub>	●							
X <sub>2</sub>	●	→	●					
X <sub>3</sub>	●	→	●		→	●		
X <sub>4</sub>	●	→	●		→	●	→	●
X <sub>5</sub>	●	→	●		→	●	→	●

Based on the considerations above, the present research examines the instances taken from the spoken texts of the Contemporary English Corpus of America (hereafter COCA) to analyze the synchronic differences in usage of the “the X is” type of expressions. The research objects are *the thing is*, *the point is*, *the trouble is*, *the problem is*, *the truth is*, *the fact is*, *the reason is*, etc.. Specifically, this article takes *the thing is* as an example to investigate its uses in different stages of grammaticalization and reconstruct its path of evolution. Special emphasis is laid on the correspondence between its form and meaning. This study further compares the relationship between the usage and form of other “the X is”-like expressions and analyzes the grammaticalization gradience within this series. This approach provides a relatively comprehensive understanding of the synchronic differences of the items concerned, particularly their role as an important source of discourse markers.

### 3. Analysis of the Usage of *the Thing is*

#### 3.1 Identifying the Relevant Item

In contemporary English, *the thing is* is variously used. It has at least two macro uses, non-parenthetical and parenthetical, which are syntactically and semantic-pragmatically different.

##### 3.1.1 Non-parenthetical *the Thing is*

Non-parenthetical *the thing is* works as a matrix clause, which can be interpreted as a grammatical “subject + verb + (complement)” construction, followed by a clausal complement. In this case, *the thing is* generally receives some stress and there is no pause between it and the subsequent sentence components.

As shown in the example below, the appearance of the complementizer *that* denotes that *the thing is* is a matrix clause and it is not syntactically detachable. Moreover, the term *the thing* here has

propositional meaning and can be rephrased as *the problem* depending on the context. Both terms are “shell nouns” and can be taken as labels of the message conveyed by the following *that*-clause (Schmid 2000).

(1) **PINSKY#** That is exactly right. Danine, how about you? Do you have any sense of what should happen with these folks?

**DANINE-MANETTE#** There is so much going on here. It is outrageous. And, the thing is that I keep hearing a lot of people talking about the mother and the mother having left the kids. ...

### 3.1.2 Parenthetical *the thing is*

Contrary to the non-parenthetical use, when *the thing is* serves as a parenthetical, it is syntactically optional, occurring “either outside the syntactic structure or loosely attached to it” (Brinton, 1996, p. 34). Observe the following example.

(2) **JOHN ROBERTS:** ... They’re not so worried about the Republicans. The thing is, with President Bush’s speech today it was a very effective speech. And he painted a picture that people should certainly be worried about.

What follows *the thing is* in this case, as illustrated in (2), is normally a declarative clause without the complementizer *that*. *The thing is* as a parenthetical is “generally marked prosodically by increased speed and lower volume” (Quirk et al., 1985, p. 1113). It forms an independent tone unit, usually reinforced by the pauses that follow or/and precede them, as well as other parenthetical items that sometimes co-occur. Notably, *the thing is* may also appear in the phonologically reduced form *thing is*, which will be examined in 3.2.3 below.

### 3.1.3 Structurally Indeterminate *the thing is*

The difference between the non-parenthetical and parenthetical uses of *the thing is* is evident in the examples discussed earlier. However, recognizing the status of *the thing is* is not always unproblematic, especially in the instances in which *the thing is* is followed by a *that*-less clause. As exhibited in (3), without formal and prosodic clues, it is not that easy to determine whether *the thing is* is an initial matrix clause or a parenthetical.

(3) **RAZ:** Anger is universal, but it’s also a complicated emotion. We tend to think of anger as an irrational response. And we think of people who get angry as unhinged. And when it’s uncontrollable, anger can be really destructive. The thing is we also need it. Anger tells us when something feels wrong or unjust or unfair. ...

## 3.2 Functions of Parenthetical *the thing is*

Parenthetical *the thing is* may serve as either a comment clause or a discourse marker.

### 3.2.1 Comment Clause Use

As presented in (4), *the thing is* as a comment clause mainly denotes the speaker’s judgment of the importance of the upcoming proposition and draws the hearer’s attention to what follows. Its comment function becomes more apparent if the evaluative adjective *important* is added before the term *thing*. *The thing is* here can be interpreted as *what is important is* or *importantly*, which clearly also serves an

emphatic and evaluative function.

(4) **MARTIN-SAVIDGE-CNN:** ... Then the next step of the process is who is going to pay for the bus ticket? Who's going to pay for the airline ticket once that is established and the money has been transferred here, wired or otherwise. Then they begin to move them off to the buses or they move them off to the airport? But the thing is, Fredericka, it's working well for now, but I asked the mayor how long can they hold it up and here's what he said. ...

*The thing is* can be used as a comment clause in a variety of contexts and can be employed when the speaker feels the need to emphasize what is to be said next, which may be an important fact, opinion, or issue worthy of consideration. In general, comment clause *the thing is* works on the level of sentence/clause, primarily evaluating the following proposition itself. Put differently, it does not link directly to previous discourse segments and its primary function is subjective (emphatic and evaluative) rather than textual like its discourse marker use.

### 3.2.2 Discourse Marker Use

When used as a discourse marker, *the thing is* is primarily textually significant, with a larger scope, operating on the level of discourse. Unlike its comment clause use, *the thing is* as a cohesive device has a clear reference to a prior discourse segment. It may signal a cohesive relation between discourse utterances (causality, contrast, topic shift, etc.) or function as a hesitation filler to indicate discourse continuation.

As shown in example (2) above, *the thing is* is used to denote a causal relation, which can be roughly replaced by *because* or *since*. *The thing is* can not only be employed to mark relations such as causality or contrast between discourse segments, but it can also function at a higher level of discourse structure, i.e. to indicate a topic shift. Clearly, *the thing is* in the latter case serves to "relate a single utterance to a property of a longer stretch of discourse" (Schourup, 1999, p. 257). Specifically, what *the thing is* triggers can be either a change of focus or perspective on the topic currently being discussed or a return to a topic that has been talked about before. A topic change may make the whole discourse lacking in coherence, and the presence of *the thing is* can make the change seem less abrupt and more acceptable, helping to strengthen the coherence of the discourse. In the following example (5) of a discussion on how to deal with spousal betrayal, one participant used *the thing is* to shift the conversation to her own personal experience.

(5) **#0:** Absolutely, I have the exact same conversation with my wife. I asked her if she would go the press conference...

**#0:428:** Right. And the thing is, and I actually went through something similar when I was married in a former life. ...

Like many other discourse markers, *the thing is* can be used as a filler, providing the speaker with time to think and indicating the continuation of utterance. As shown in example (6), *the thing is* is unique as a filler in its frequent co-occurrence with other discourse markers.

(6) **BUTEAU#** No, you could actually just live your life, and be quiet, and you know, have like a

normal childhood. I don't understand it at all. This is definitely just a paycheck so they can do something else this year or next year. I mean, the thing is, like, you know, Jon Gosselin is on VH-1's celebrity therapy, couples' therapy. ...

### 3.2.3 Variant of *the Thing is*—*thing is*

*The thing is* may also occur in the phonologically reduced form *thing is*, but such examples are not commonly found in COCA (12 occurrences). *Thing is* serves as a parenthetical and does not have a non-parenthetical counterpart as *the thing is*. In fact, there are no instances in which *thing is* is followed by a *that*-clause with an explicit complementizer. Prosodically, a distinct pause (embodied in writing as a comma) invariably occurs between *thing is* and its following sentence constituents and as a result, independent tone units are formed. It seems that the omission of *the* makes the construction a more particle-like unanalyzable unit. Moreover, with the exception of one example of comment clause use, *thing is* in my data is used as a discourse marker in the other 11 instances, denoting a contrastive discourse relation, a topic shift or discourse continuation.

*Thing is* in the example below is used as a hesitation filler, and its presence provides the speaker with thinking time during which he/she can formulate what is to come. Like *the thing is*, *thing is* on such occasions also tends to co-occur with other discourse markers.

(7) **SARA-HAINES#** Very Freudian (inaudible). Beyond. I don't really picture, like, Hillary Clinton, let's say, or Margaret Thatcher wanting to be tied up, I don't see it. You see, I disagree. Do you? I do. You do? I really do. You know, thing is, look, I'm a dominant, we're both, fairly dominant people. We're comedians. especially I'm tall, so a lot of guys are, like, intimidated by me. ...

### 3.2.4 Grammaticalization of *the Thing is*

Based on the different uses of *the thing is* and its variant *thing is*, it may be safe to presume that the discourse marker (*the*) *thing is* may stem from *the thing is* as a matrix clause, and it may follow the path of “matrix clause > indeterminate: matrix or parenthetical > comment clause > discourse marker” in the process of grammaticalization, undergoing decategorialization, desemanticization, pragmatic strengthening, phonological reduction (in the case of *thing is*), increased scope of modification and subjectification, and ultimately developing into a discourse marker.

The formulaic parenthetical *the thing is* probably originates from the construction of “the thing is + that- clause”, in which the complementizer *that* clearly indicates its status of matrix clause. If *the thing is* were to be removed, the original sentence would become incomplete and ungrammatical. The shift of the structure from a matrix clause to a parenthetical results from the omission of the complementizer *that*. In the construction of “the X is+ that-less clause”, the status of *the thing is* is indeterminate, either a matrix clause or a parenthetical. In fact, it is the structural indeterminacy that allows a reversal in syntactic hierarchy (Brinton, 2008, p. 246), elevating the original dependent *that*-clause into the matrix clause and downgrading the original matrix clause to a parenthetical loosely attached to the syntactic structure. *The thing is* finally becomes a true parenthetical as a result of further decategorialization.

In the evolution of *the thing is*, semantic-pragmatic change and syntactic reanalysis occur

simultaneously. In other words, while the construction is decategorized from a matrix clause to a formulaic particle-like parenthetical, it also undergoes desemanticization and subjectification. Specifically, the head noun *thing* loses its concrete propositional meaning and the construction takes on a more abstract meaning, conveying the speaker's emphasis of the importance of the proposition that follows. What is also involved in this process is increased scope of modification and syntactic freedom. As a result, *the thing is* develops into a comment clause that operates at the sentence/clause level, primarily expressing the speaker's subjective attitude towards the proposition.

Discourse marker *the thing is* is the result of further grammaticalization of the structure, or rather an advanced stage of its grammaticalization process. Comment clause *the thing is* is further desemanticized and pragmatically enriched. With an even larger scope, it operates at the discourse level and its primary function is no longer to express the speaker's subjective evaluation, but to serve as a textual device promoting discourse coherence. Additionally, the item concerned undergoes phonological attrition or erosion in this process, sometimes appearing in the form of *thing is*.

The various discourse marker functions of *the thing is* can be regarded as the generalized and conventionalized conversational implicatures of the item as a comment clause through repeated use in different contexts (Cf. Aijmer & Simon-Vandenberg, 2004). It should be noted that when *the thing is* is used as a discourse marker, its evaluation and emphasis are greatly weakened, although we cannot completely deny the existence of its subjective meaning. Therefore, *the thing is* in example (2) can no longer be replaced by *the important thing is*, *what is important is* or *importantly*; instead, some other discourse markers that can denote causality, such as *because* or *since*, can be used to accurately explain *the thing is*. Obviously, the two parenthetical uses of *the thing is* are both related and distinct: its discourse marker function can be understood as developing from, but going beyond its comment clause use.

To sum up, the evolution of *the thing is* from a matrix clause to a discourse marker is: [*the thing is* + *that*-clause (matrix clause)]→[*the thing is*+ *that*-less clause (matrix clause or comment clause)]→[*the thing is*, *that*-less clause (comment clause)]→[(*the*) *thing is*, *that*-less clause (discourse marker)], which clearly demonstrates the relationship between the different functions and forms of *the thing is*.

#### 4. Other “the X is”-type Constructions and the Grammaticalization Gradience

##### 4.1 Other constructions

In addition to *the thing is*, other “the X is” type of expressions with epistemic meaning (e.g., *the fact is*, *the problem is*, etc.) also have both non-parenthetical and parenthetical uses. In the following examples (8), (9) and (10), *the fact is* is used as a matrix clause, a comment clause and a discourse marker respectively:

(8) **Ms. HUSTVEDT:** Usually, the fact is that when I'm in pain, almost every kind of sensory stimulus is painful....

(9) **MITCHELL:** ... She's not an American, but the fact is, it's kind of a hard song to sing and a

difficult song to put across. And she did it beautifully.

(10) **GILLESPIE:** But look, we are not interested in attacking Albert Gore. That is not the cause here. The fact is, we want to talk about Governor Bush, and his strong leadership style...

There are also instances, as in example (11), in which *the fact is* is structurally indeterminate.

(11) **PINSKY:** ... We don't know if there's other thing as well. And she may have just been manic and psychotic. And the fact is I don't know—first, ...

Similar to *the thing is*, *the fact is* as a discourse marker can also be used as a cohesive device to mark contradictory relation, topic shift, etc., or as a filler to indicate discourse continuation. In example (10) *the fact is* is used to lead to a topic change.

Moreover, as illustrated in the example below, the phonologically reduced form *fact is* of *the fact is* also appears in my data. In a total of 45 instances, *fact is* is used as either a comment clause or a discourse marker. Actually, all of the head nouns with the exception of *reason* have a certain number of “X is”-type instances.

(12) **Mr-HARRIS:** The place is not for sale it's what I like to say. But if somebody offers me what I think is way, way, way, way too much money, I'd be foolish not to consider it. Fact is, they're making people a lot faster and they're making this land and it won't be today or tomorrow, but sooner or later.

Obviously, *the thing is* and *the fact is* have similarity in usage, but do they differ in the distribution of their functions as “the X<sub>4</sub> is” and “the X<sub>5</sub> is” in Table 1? Similarly, what is the case of *the trouble is*, *the problem is*, etc.? Additionally, previous studies have suggested that many other expressions that fit this structure, such as *the reason is* and *the difference is*, cannot be used as discourse markers. What, then, are the main uses of such expressions? Based on the theoretical analysis above, the following section will examine the items concerned from the perspective of the degree of grammaticalization.

#### 4.2 The Grammaticalization Gradience of “the X is”

Discourse markers are the result of grammaticalization, so whether an expression can develop into a discourse marker can be seen as an important indicator of its degree of grammaticalization. Although there are differences in previous studies regarding the characteristics and functions of discourse markers (e.g., Brinton, 1996; Schourup, 1999), in general, the criteria for measuring whether an expression has evolved into a discourse marker include: (a) from a relatively stable linear position syntactically to the absence of positional constraints; (b) from having a well-defined function and status syntactically to having no syntactic function and being optional; (c) from expressing a specific propositional meaning to having a textual or interpersonal function; and (d) from a metrically non-independent form to forming an independent tone unit. The four criteria above correspond to flexibility of position, transition from non-parenthetical to parenthetical, functional distribution of parentheticals, and formation of independent tone units. Among them, the formation of independent tone units can be reflected in the formation of parentheticals. Therefore, we can examine the degree of grammaticalization of “the X is”-type expressions and related issues from three other perspectives. Based on different head nouns, namely, *thing*, *point*, *trouble*, *problem*, *truth*, *fact*, *reason*, this study



compares and analyzes all the obtained instances of the “the X is” type of expressions as matrix clauses, indeterminate usage and parentheticals in the spoken data. The results are as follows.

#### 4.2.1 Distribution of non-parenthetical and Parenthetical Uses

Table 2 shows that the distribution of different uses of each expression varies significantly. The parenthetical use of *the thing is* is 2.78 times more frequent than its matrix clause use. It is the highest among the expressions in the category of “the X is”, followed by *the truth is*, and the lowest is *the reason is*. *The fact is* has the highest and similar numbers of the two macro uses among the items concerned, but notably, the frequency of the matrix clause use of *the fact is* is nearly 7 times that of *the thing is*, while its parenthetical use is only 2.33 times that of *the thing is*. In comparison, *the thing is* is probably more grammaticalized. *The reason is* has the lowest frequency of occurrence in the form of a parenthetical, which is about one-third of its matrix clause use. Therefore, from the distribution of different uses, the grammaticalization gradience of the “the X is”-type expressions may be presented from high to low as follows: “thing > truth > trouble > point > fact > problem > reason”.

**Table 2. Distribution of Different Uses of “the X is”**

X	Non-parenthetical	Parenthetical	Parenthetical/non-parenthetical	Indeterminate
thing	211	587	2.78	467
truth	363	632	1.74	223
trouble	57	93	1.63	101
point	511	549	1.07	245
fact	1437	1368	0.95	804
problem	1156	884	0.76	1623
reason	168	57	0.34	294

#### 4.2.2 Characteristics of “the X is” as a Parenthetical

The difference in the degree of grammaticalization of the “the X is”-type expressions is also reflected in the distribution of the specific functions of the relevant items used as parentheticals. A certain number of instances were randomly selected from the data of parentheticals mentioned above to analyze their specific usage, namely, comment clause, discourse marker and indeterminate use (see Table 3). It is found that *the thing is* has the highest percentage (57%) as a discourse marker. *The fact is* and *the truth is* have a significantly lower percentage of discourse marker use compared with *the thing is*. In the instances of *the trouble is* and *the problem is*, there are a large number of cases of “dual identity” with both comment clause and discourse marker function, which can be interpreted as *but the trouble is/the bad thing is/the unfortunate thing is/the regrettable thing is/the crucial thing is/...*, while the proportion of the instances used solely as discourse markers has decreased obviously. *The point is*, on the other hand, is mostly used as a comment clause and has an emphatic meaning. This suggests that

*the trouble is*, *the problem is* and *the point is* are not highly grammaticalized. In addition, no discourse marker use was found in the parenthetical form of *the reason is* retrieved from the corpus. In some instances *the reason is* is emphatic to a certain extent, but it cannot yet be regarded as a comment clause in the strict sense due to its clear propositional meaning. Therefore, in terms of the distribution of parenthetical uses, the grammaticalization gradience of the “the X is”-type expressions, from high to low, is: “thing > fact > truth > trouble > problem > point > reason”.

**Table 3. Distribution of Different Functions of Parenthetical “the X is”**

X	Discourse marker	Comment clause	Indeterminate	Total
thing	114	78	88	200
fact	172	208	20	400
truth	82	112	6	200
trouble	16	16	18	50
problem	87	135	78	300
point	56	134	10	200

#### 4.2.3 Syntactic Position

Next comes the syntactic position, i.e., sentence-initially, sentence-medially or sentence-finally, of all the instances of “the X is” type of expressions as parentheticals in the spoken data. Theoretically, parenthetical expressions can occur in various positions as a direct result of their loose connection with the syntactic structure of the sentence. Moreover, when the objects in question appear in non-initial positions, they are clearly not integrated into the syntactic structure and can therefore be unquestionably regarded as parentheticals.

The results show that there are no instances in which the items under study appear sentence-finally. In terms of the proportion of sentence-medial occurrences, *the thing is* is the highest, indicating its comparatively flexible syntactic position and possibly the highest degree of grammaticalization. The proportions of *the fact is*, *the truth is*, *the point is* and *the problem is* are relatively close and significantly lower than that of *the thing is*. *The trouble is* and *the reason is* are even lower. Roughly speaking, from the perspective of sentence-medial position, the degree of grammaticalization of “the X is” type of expressions is, in descending order, “thing > fact > truth > point > problem > trouble > reason”. At the same time, it can also be seen that, in general, the number of non-sentence-initial positions of such expressions is not large and their syntactic distribution is not flexible enough. This may be related to their source structure (matrix clause), and also indicates the degree of grammaticalization of some of them is not high enough.

**Table 4 Syntactic Position of Parenthetical “the X is”**

X	sentence-initially	sentence-medially	percentage of sentence-medial position (%)
thing	457	130	22.15
fact	1201	167	12.21
truth	561	71	11.23
point	490	59	10.74
problem	790	94	10.63
trouble	90	3	3.23
reason	56	1	1.75

## 5. Conclusion

Based on the above analysis, it can be seen that although the “the X is” structure is an important source of English discourse markers, the degree of grammaticalization of the expressions that conform to this structure varies, basically forming a gradience of “the thing is > the truth is/the fact is > the trouble is/the point is/the problem is > the reason is/...”. Among them, *the thing is* has the highest degree of grammaticalization and is often used as a discourse marker; *the truth is* and *the fact is* have a lower degree of grammaticalization than *the thing is* and are used similarly; *the trouble is*, *the problem is* and *the point is* are basically on the same gradience and are much less grammaticalized, mostly used as comment clauses with a more prominent subjective evaluative function. Notably, semantically similar expressions in the source structure, such as *the truth is* and *the fact is*, as well as *the trouble is* and *the problem is*, not only share a comparable degree of grammaticalization, but also have similar uses. For example, both *the trouble is* and *the problem is* have prominent negative evaluative functions.

The reason for presenting the above gradience may be related to the propositional meaning of the head noun “X” of the source structure. For instance, the propositional meaning of *thing* itself is more abstract and general, making it easier to bleach or even disappear, transforming into a subjective meaning that indicates emphasis. This further evolves the expression into a discourse marker. Other members of the “the X is” construction in which “X” is not epistemic are the least grammaticalized and the evolution of their semantic-pragmatic functions may lag behind that of the syntactic structure. For instance, although *the reason is* may appear in the form of a parenthetical, and even therefore shows some flexibility in syntactic position, it is semantically identical to *the reason is* as a matrix clause. In other words, in the process of its syntactic reanalysis, desemanticization and subjectification are not completed, and thus it has not yet been able to evolve into a full-fledged comment clause or even further a discourse marker. Furthermore, we did not find any “X” that only has the matrix clause use. Put differently, “X” that can be used in this construction has at least two uses: “the X is+that-clause” and “the X is+that-less clause”. The flexibility of whether the complementizer *that* is missing or not indicates that this type of expression is easy to undergo syntactic reanalysis, thereby further

grammaticalization.

Of course, this study is preliminary and there are still some issues that have not been examined. For example, what are the differences in the usage of related members of series-occurring discourse markers with semantically similar source structure? Does the “the X is” construction have a cross-linguistic commonality with similar structures in other languages, such as *wentishi* in Chinese? These issues still require further research in the future.

## References

- Aijmer, K. (2007). The interface between discourse and grammar: *The fact is that*. In A. Celle, & R. Huart (Eds.), *Connectives as discourse landmarks* (pp. 31-46). Amsterdam/Philadelphia: John Benjamins.
- Aijmer, K., & Simon-Vandenberg, A. M. (2004). A model and a methodology for the study of pragmatic markers: The semantic field of expectation. *Journal of Pragmatics*, 36(10), 1781-1805.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.
- Brems, L., & Davidse, K. (2010). The grammaticalization of nominal type noun constructions with *kind/sort of*: chronology and paths of change. *English Studies*, 91(2), 180-202.
- Brinton, L. J. (2008). *The comment clause in English*. Cambridge: Cambridge University Press.
- Brinton, L. J. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin and New York: Mouton de Gruyter.
- Cao, X. L. (2010). From subject-predicate structure to discourse marker—Grammaticalization of “wo/ni V”. *Chinese Language Learning*, (5), 38-50.
- Crystal, D., & Davy, D. (1975). *Advanced conversational English*. London: Longman.
- Prévost, S. (2011). *A propos* from verbal complement to discourse marker: A case of grammaticalization? *Linguistics*, 49(2), 391-413.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Schmid, H. J. (2000). *English abstract nouns as conceptual shells*. Berlin and New York: Mouton de Gruyter.
- Schourup, L. C. (1999). Discourse markers. *Lingua*, 107(3-4), 227-265.
- Thompson, S. A., & Mulac, A. (1991). The discourse conditions for the use of the complementizer *that* in conversational English. *Journal of Pragmatics*, 15(3), 237-251.
- Traugott, E. C. (1995). *The role of discourse markers in a theory of grammaticalization*. Paper presented at the 12th International Conference on Historical Linguistics, Manchester, UK.
- Wang, H. X. (2014). Functional commonalities and differences of English discourse markers—A study based on the persistence principle of grammaticalization. *Foreign Language Teaching and Research*, 16(5), 691-703.