Truth Is, Sentence-Initial Shell Nouns Are Showing Up Bare

Laurel Smith Stvan

University of Texas at Arlington, Texas, USA
stvan@uta.edu

Abstract: In one subtype of English shell noun construction, the noun serves as the subject in a pre-clausal unit, e.g., “The thing is.” Shell noun NPs have mainly been investigated synchronically, often as case studies of particular noun types, with the bare forms only ever briefly noted. Data from COCA and COHA was examined to collect the range of sentence-initial bare form shell nouns, to track any changes in their use occurring between 1810 and 2012. The findings suggest that, referentially, these abstract bare forms function differently than bare concrete count nouns, and distributionally, that bare shell forms are used increasingly in subject position, confirming their state as grammaticalizing discourse marker constructions.

Keywords: bare nouns; shell nouns; discourse markers; diachronic English; grammaticalization.

1. Introduction

Recent work has established a set of English nouns which, while normally found as count nouns, also occur lacking both articles and plural forms in a non-mass, non-count use that has the distribution of a full NP (e.g., she was in prison, they ate at home, we left school, camp was exhausting). The most frequent uses for these bare singular count nouns (BSCNs) occur when they are in PPs, where they may be used to refer to individuals or kinds (Stvan 2007), or to capacities (de Swart et al. 2007). These contrast with other restricted semantic sets of bare singulars that occur as predicates, and not as arguments at all.

In a number of languages, including English, a quite limited set of BSCNs appear in subject position (Stvan 1998). A systematic look at which bare singular forms do show up in subject position, and the ways that these words are interpreted, is now more easily
possible due to larger corpora that are tagged for part of speech (POS), though such searches are still hampered by the fact that bare singulars are not an identified subset; most POS tagsets allow a search for “common nouns that are not plural,” but these do not separate out the vast number of mass nouns from the more marked set of bare singular NPs.

While a subset of bare concrete location nouns shows up as subjects (e.g., *home is where the heart is, school kept them busy that week*) this paper focuses on another set of bare subject nouns—constructions with bare abstract count nouns in pre-clausal position. Often taken to be colloquial variants, typical instances include *Truth is, Problem is,* and *Trouble is.* This paper opens up the set of abstract nouns that are subjects in pre-clausal constructions to look at those that lack articles. The goal is to show what the behavior and frequency of these bare forms might reveal.

Just as BSCNs are most often found in PP constituents, these abstract bare forms are frequently examined as subjects in a clause-like unit that itself precedes another full clause, as seen in the examples in (1). The main “complement clause” can be a statement, as in the underlined sections of (1a) and (1b) or a question, as underlined in (1c) and (1d); while the pre-clausal unit provides the kinds of evidential, attitudinal, and meta-talk information often conveyed by discourse markers (Schiffrin 1987).

(1) (a) “Truth is, he’s never seen anybody bust their head falling off a bike, but he’s seen a few people crash.” [COCA:*The Antioch Review,* 2009]

(b) “They aren’t real monks, anybody can tell that. Trouble is, our robes and our rules of silence make anonymity so easy here, and monastic discipline makes questioning authority so hard.”
[COCA:*Fantasy & Science Fiction,* 2007]

(c) “Question is, will they make history by making space tourism possible?”
[COCA:*Anderson Cooper 360°,* 2004]

(d) “Yeah, yeah. Thing is, how’d you figure it out?” [COCA:*Analog,* 1999]

Keeping these kinds of structures in mind, the paper addresses the topic in the following ways: Section 2 offers a look at relevant previous literature on bare nouns, shell nouns, and discourse markers. Section 3 looks at these noun forms, focusing on their position with sentences, offering, in particular, some motivation for looking at shell nouns as subjects, a discussion of whether these functions vary if the noun phrase has an article or is bare, and then an analysis of the referential uses of these constructions. Section 4 presents a look at the corpus methods used to obtain the data. The data is analyzed in Section 5, for both synchronic and diachronic corpora. I conclude by detailing implications that follow from a change in frequency of bare form subjects.
2. Literature Review

A number of aspects of both shell noun and bare noun behavior have been looked at in the past 15 years. The current findings intersect with the following areas of research: work on shell nouns: Schmid (2000), Delahunty (2011; 2012); on bare singular count nouns: Stvan (1998; 2007); on bare NPs as referring expressions: Carlson and Sussman (2005); Kolhatkar et al. (2013); and on these forms as types of pragmatic markers: Aijmer (2007), Günthner (2007), Brinton (2010), and Keizer (2013). I pull these together below to describe two aspects of these constructions: their grammatical position and their discourse function.

2.1 On Shell NPs in Subject Position

The nouns under discussion have abstract referents (thing, problem, issue), leading Schmid (2000) to call them “shell nouns”—words that do not have much semantic content, serving instead as shells for a type of information. In this paper I focus on one template in which shell nouns occur, as subject of a copula forming a pre-clausal unit. As subjects, of course, these are actually noun phrases, not nouns, regardless of whether they appear with an article, a modifier, or are completely bare.

Distributionally, these pre-clausal units are similar to a subset of discourse markers that Brinton (2010, 285) refers to as “abridged clauses,” which includes items like you know, I mean, or you see. These truth is type clauses, however, differ from the abridged discourse marker clauses in that they cannot occur throughout a sentence, but are found only pre-sentence initially, and, more particularly, are used to introduce a second clause, sometimes with an optional complementizer. (Though see Günthner [2007] on characteristics that, in German versions, make the two components not typical of the “matrix clause” and “subordinate clause” labels.) The template variations for this shell noun construction are shown in (2), i.e., there is an optional article, optional comma, and optional complementizer.

(2) (the) truth is (,) (that) [you didn’t show up]

Delahunty (2012) calls these units “thing sentences.” He mentions the small number of bare forms that are found, concluding that this is essentially a definite NP form. In particular, he cites his earlier work, which found that “97% of . . . examples were definite,” and work by Brenier and Michaelis (2005), who found that “so many of their examples were definite that they concluded that definiteness is a defining characteristic of the construction.” He also notes Brenier and Michaelis’s claim that the forms “are highly lexically restricted, [and] have an invariant morphological form (they contain the definite article)” (Delahunty 2012, 60). In that explanation, however, it seems that definiteness is assumed by the presence of a determiner, although recent work in nominal expressions has noted the mismatch between some NP forms and their referent’s identifiability (Carlson and Sussman 2005; Stvan 1998).
Regarding the bare shell noun forms, Delahunty notes that “Thing sentences allow a number of elisions that are associated with factors such as style, register, and mode” which allow the definite article to be dropped, but that “their meanings and discourse functions are identical to those of full TSs” (Delahunty 2012, 42). The current paper calls into question the claim that the bare forms are identical in discourse function to the articulated version.

2.2 On DMs as Grammaticalized Items

These pre-clausal units, do, however, fit several aspects of the general description of discourse markers summarized by Brinton (2010). For example, they “preferentially occur in sentence-initial position. They are syntactically independent elements that are loosely attached to (parenthetical to) their host clause and often constitute a separate intonation unit . . . They occur with high frequency, especially in oral discourse, and may be stylistically stigmatized” (Brinton 2010, 285). Indeed, stylistically, at first glance they have a “homey” or “colloquial” sense. But what more formally characterizes the nouns used in this way?

This set of [abstract count noun + verb] units has been discussed specifically and extensively by Schmid (2000) as a sub-pattern of shell noun use. Their position before a clause led Biber et al. (1999) to call the clausal units “utterance launchers.” Functionally, Günthner (2007, 6) categorizes them as projector-phrases, “a rhetorical deferral of the focal” to the following material; she notes that this idea “lines up with certain tendencies in grammaticalization” (2007, 15).

They have been framed as a kind of text deixis or “discourse label” that names or announces the format of the larger entity being discussed (Francis 1994). A variation of this construction containing a definite noun followed by two copula forms (“the thing is is”) was analyzed by Tuggy (1996), created where an NP plus copula merges with a focus formula, though I will not explore these here.

In examining four particular NP types in pre-clausal NP + copula forms from BNC data, Aijmer (2007, 39–40) also suggests that these forms are undergoing grammaticalization, whereby the unit is moving from a traditional matrix clause to a position as a pre-front field constituent, and on to a pragmatic marker, during which the semantic meaning is bleached, and the unit is pragmatically enriched. Keizer (2013) also notes that this construction serves as a pragmatic-rhetorical marker with a presentative function. All these authors emphasize the marked discourse function of the form, but focus primarily on the articulated noun and a synchronic description of the construction’s use.

2.3 The Present Research Questions

It is clear that a number of questions have begun to be asked about these bare shell noun forms. The present work started out with a set of descriptive issues, as shown in (3).
(3) (a) What is the range of forms used in these constructions?

(b) Do the nouns share any lexical semantic features?

(c) Do the phrases share any discourse uses?

(d) What words are excluded from this position?

(e) Are they all possible with articles, too?

Several of these issues were briefly examined above. But in particular, this paper will look at the following more focused research questions:

RQ 1: Are these nouns used referentially?

1a) Are they any less referential when lacking an article?

RQ 2: Is there evidence of diachronic movement towards bare truth is forms?

2a) Is it happening any more across time?

2b) And, if so, does that signal grammaticalization?

In short, I am specifically interested in the smaller percentage of uses when the article of the pre-clausal NP can be deleted, in whether this functions semantically or pragmatically any differently than the version with a full NP, and whether this ratio of bare to full is changing. I will start by taking a look at some differences that show up based on the distribution of bare shell nouns within a clause. These sentence types are often examined as a semi-fixed construction. One piece of evidence of their template-ness is that shell nouns in other grammatical positions cannot show up bare in the same way.

3. Grammatical Position of Bare Forms

3.1 Distribution of the Shell NPs

First of all, we can see that variants of [Noun is] sentences can also show up with the shell noun occurring in a more rightward position in a pre-clausal unit, as in the pre-clausal construction Here’s the thing. However, when the shell noun is in the position after is, the article cannot be dropped. While example (4) shows that in the pre-clausal position the article is optional for the subject nouns, this is not the case when the noun occurs after the verb. It is ungrammatical to begin with Here’s thing, and likewise with
other such nouns, as shown in (5). So some of the ability of the nouns to show up bare has to do with their position in the sentence, or their slot in the construction.

(4) (a) The thing is, she’s not home. pre-copula + article
(b) [ ] Thing is, she’s not home. pre-copula – article

(5) (a) Here’s the thing, she’s not home. post-copula + article
(b) *Here’s [ ] thing, she’s not home. post-copula – article
*Here’s [ ] problem, she’s not home. post-copula – article
*Here’s [ ] issue, she’s not home. post-copula – article
*Here’s [ ] deal, she’s not home post-copula – article

As noted earlier, tagged corpora make it easier to focus on gathering subsets of examples of lexical categories from texts, but separating mass and count, let alone smaller subsets such as bare singular count nouns, is still tricky. Such is the case with (COCA) the Corpus of Contemporary American English, which uses the Claws7 tagset. Initial searches in COCA for a non-plural noun followed by the word is bring up mainly mass nouns, gerunds, and names, in the noun slot. (Top subject hits are life, admission, cost, problem, truth, congress, it, love). Also, not all readings of English nouns have distinct morphosyntactic indicators. So there can be false positives involving identical looking mass and count nouns. For example, in (6) we can see contrasts with the noun truth:

(6) (a) Truth is stranger than fiction
[mass, referential subject]
(b) All truths are easy to understand once they are discovered
[count, referential subject]
(c) The truth is, he didn’t do it.
[count, pragmatic unit, referential?]
(d) Truth is, he didn’t do it.
[count, pragmatic unit, referential?]

The word truth can be countable and can also be used as a mass noun. In (6a) we can see that the combination truth is can show up as a mass noun used referentially, where it takes a predicate; in (6b) truth is a countable noun used referentially. In the constructions of
interest, in (6c) and (6d) we see a freestanding phrasal unit that introduces a clause, which can be used with both articulated and bare forms.

I suggest that the (6c) and (6d) forms are countable, with (6d) representing an NP with a deleted article. But are the NPs in (6c) or (6d) referential? One test is to check for their use in discourse anaphora. For example, the truths in (6b) can be referred back to with the pronoun they. But with the pre-clausal uses in (6c) or (6d), this is more problematic.

3.2 Occurrence in Discourse Anaphora

Further illustration of their referential qualities can be seen by examining the constructions with regard to discourse anaphora.

(7)  (a) The truth is, we don’t have enough cash. #It is a big one.

(b) “The trip was meant to encourage Syria along the path of peace. Problem is, it didn’t.” [COCA:Time, 1994]

(c) The truth is that you’re never going to know.

(d) The truth is, which is a fact you’ll have to acknowledge, that you’re never going to know.

In (7a), direct co-reference with a pronoun does not work for the problem. In (7b), an example from the corpus, the word it shows up in the following sentence, but not as co-referential with problem; it co-refers with an NP from earlier in the discourse. In (7c), we see that what these shell nouns are co-referential with is the entire clause that they introduce, a unit they foreshadow and name. Thus, the pre-clausal shell nouns have a kind of text deixis function where the referent is the upcoming clause. And in (7d), we see that the shell noun can be co-referential with a restatement that is an elaborately modified NP, as well as with the clause it is introducing. So while pronouns do not work well here in anaphora, shell nouns can be referred back to by later recaps or syntactic elaborations of the clause.

However, this seems partly sensitive to whether the noun and verb of the pre-clausal units are kept together or divided by the follow-up clauses. This can be seen by examining instances of non-splittable antecedents in (8).

(8)  (a) The truth—and it’s a doozy—is that you’ll need to eat cookies every day.

(b) *Truth—and it’s a doozy—is that you’ll need to eat cookies every day.
(c) “Thing is, and I have learned this from working at the—the, with the tribal people, the Coeur d’Alene people, I never understood how important it was to know where you came from.” [COCA: Larry King, 2001]

(d) “Truth is, though, and you know it at sight and without a second thought, Barfoot has known every kind of pain.” [COCA: Virginia Quarterly, 1990]

In (8a), the articulated form can be referred back to with a pronoun, while in (8b), the bare form cannot. However, there is also a contrast between (8b) and (8c), where both forms start with bare nouns. But with (8c), the pronoun occurs after the completion of the pre-clausal truth is unit.

It is not, however, just the intactness of the pre-clausal unit that limits the use of following pronouns. Their further discourse anaphoric abilities via later pronouns are quite rare: only three examples showed up in in COCA, shown in examples (8c), (8d), and one other. This suggests a limited and very different referential ability than that shown by concrete bare singular count nouns (Stvan 2007).

One aspect not tracked in these short text excerpts is cases where the article is called for due to previous mentions of the referent. (One such tactic is explored by Kolhatkar et al. [2013] who present computational methods to find such earlier mentioned referents of NPs made up of anaphoric shell nouns such as this issue. Their method is to reverse engineer the hunt, focusing on cataphoric shell NPs which have immediately following antecedents, e.g., The fact that x to use as training data to annotate future discourse anaphora.)

Bare forms also show up with pre- and post-modifiers, indicating that the truth is type construction is not a question of light vs. heavy NPs. For certain nouns, modification also interacts with ability to lack an article.

3.3 Bareness and Modification of These Shell Noun NPs

Aijmer (2007, 33) notes two aspects of fact is constructions in British English: when the complementizer “that is present the definite article must also be present. There were no examples such as fact is that (a single example was found in the whole BNC) . . . When the is missing, fact is not usually preceded by an adjective (*Simple fact is, *simple fact is that).”

In American English, the bare form collocation of fact is that also shows up low in frequency; there were only two examples found in COCA. And no examples show up of unarticulated fact preceded by an adjective. Others of these shell nouns, however, showed up many times with adjectives, as either bare forms + that or as sentence-initial bare forms.

Bare forms show up with post-modifiers, as typified in (9a–c), and with both pre- and post- modification, as in (9d).
(9) (a) “**Thing** I like about Stairmaster is, it requires only about the minimum of 20 minutes, and you can—in and out, and you’re off and starting your day. That’s the thing I like.” [COCA: *Morning Workouts*, 1998]

(b) “Hi. Great show. **Question** I haven’t heard on the show yet is about the effect of demographics on the market.” [COCA: *Talk of the Nation*, 2000]

(c) “**Problem** with that is, it might be Bush’s war, but it’s General Piraeus’s strategy.” [COCA: *Fox News All-Stars*, 2007]

(d) “**Great thing** about this site is you can get clothing—designer clothing at amazing discounts.” [COCA: *CBS Morning*, 1998]

To explore bare form distribution, and knowing that bare forms in particular occur less often, I mined two corpora of American English for examples. These sources are detailed in the Methodology section below.

### 4. Methodology and Data Sources

I queried two online corpora of American English: the Corpus of Contemporary American English (COCA), containing 450 million words, from texts from 1990–2012, and the Corpus of Historical American English (COHA), composed of texts from 1810–2009, which contains 400 million words in total, but contains fewer sources in the older sections. Tokens were found by exploiting the built-in POS tags, such as the query illustrated in (10). This query asks for all instances of punctuation (so that what follows a comma, semi-colon, period, or quotation mark will include clause-initial forms), followed by a non-plural noun, followed by all forms of the word *be*, followed by another piece of punctuation (a comma, a period, a colon, a dash, etc.).

(10) [y*] [*nn1*] [vb*] [y*]

I also queried with the final element replaced with a symbol for conjunction [e*], to gather examples where no punctuation separates the two components, but instead, the second clause starts with a complementizer such as *that*, *whether*, *if*, etc.

### 5. Analysis of the Synchronic Corpus Data

#### 5.1 Range of Be Forms Found

The range of inflectional forms of the word *be* in the pre-clausal units was examined, since in studies of particular nouns, other researchers have noted that these copular forms are limited primarily to simple forms (e.g., Kolhatkar et al. [2013, 302]):
“they are generally expressed in the present tense”; Aimer [2007, 32]: the formula “has a fairly fixed form with a copula in the present tense”; Keizer [2013, 291]: the construction “only allows simple tenses, and in the large majority appears in the present”). In the corpus, the copula did show up more often followed by present, rather than past or participle forms of be, though not exclusively so. The outcome of searching for simple tenses can be seen in Table 1. (More complex tenses, however, can also be found, e.g., where clauses follow the problem’s been, the problem will be, the problem is going to be, the problem could be, and the problem would be.)

<table>
<thead>
<tr>
<th>Bare N + is</th>
<th>Bare N + was</th>
<th>Bare N + being</th>
</tr>
</thead>
<tbody>
<tr>
<td>problem</td>
<td>296</td>
<td>problem</td>
</tr>
<tr>
<td>truth</td>
<td>285</td>
<td>trouble</td>
</tr>
<tr>
<td>trouble</td>
<td>275</td>
<td>time</td>
</tr>
<tr>
<td>fact</td>
<td>207</td>
<td>truth</td>
</tr>
<tr>
<td>thing</td>
<td>131</td>
<td>word</td>
</tr>
<tr>
<td>word</td>
<td>116</td>
<td>thing</td>
</tr>
<tr>
<td>question</td>
<td>48</td>
<td>fact</td>
</tr>
<tr>
<td>point</td>
<td>42</td>
<td>rumor</td>
</tr>
<tr>
<td>rumor</td>
<td>18</td>
<td>question</td>
</tr>
<tr>
<td>reality</td>
<td>7</td>
<td>point</td>
</tr>
<tr>
<td>reason</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>speculation</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Inflections of be found in “problem is” type constructions.

Table 1 shows that while the majority of the cases use is, a fair number use the past tense. The smallest number use being, which can be parsed not so much as a progressive form of the copula, but rather as functioning like other participle forms that precede full clauses to convey causative conjunctive meaning. As such, they resemble constructions like those with granted and given, as shown in (11c) and (11d).

(11) (a) And you, you, you can bolster that argument in many, many ways. **Point being** we are living at a time when society is the most complicated, interconnected, immediate we’ve ever seen. [COCA::Meet the Press, 2011]

(b) I want to make sure our tee shot goes past Tiger’s, **reason being**, Tiger has a certain reputation—deserved, deserved-for putting on these shows on the last hole [COCA::Sports Illustrated, 2000]
(c) Strangely, it didn’t hurt at all—perhaps not so strangely, granted that she was swaddled in armor. [COCA: Queen of Candescence, 2007]

(d) It seems like it would be a natural, given it’s the one thing that the speaker hasn’t addressed. [COCA: All Things Considered, 1995]

5.2 High and Low Frequency Nouns Found
Table 1 showed the numbers for each COHA and COCA hit for the most frequent tokens occurring with the present tense, those with five or more hits. For a look at the variety of shell nouns that shows up with this form, we might take a look at the tail, down to those with just one hit, as shown in Table 2.

<table>
<thead>
<tr>
<th>Bare N + is</th>
<th>Bare N + is</th>
<th>Bare N + is</th>
</tr>
</thead>
<tbody>
<tr>
<td>deal</td>
<td>conclusion</td>
<td>rule</td>
</tr>
<tr>
<td>story</td>
<td>controversy</td>
<td>secret</td>
</tr>
<tr>
<td>challenge</td>
<td>danger</td>
<td>talk</td>
</tr>
<tr>
<td>consensus</td>
<td>difference</td>
<td>theory</td>
</tr>
<tr>
<td>explanation</td>
<td>downside</td>
<td>thinking</td>
</tr>
<tr>
<td>legend</td>
<td>effect</td>
<td>tradition</td>
</tr>
<tr>
<td>part</td>
<td>evidence</td>
<td>translation</td>
</tr>
<tr>
<td>pity</td>
<td>history</td>
<td>trick</td>
</tr>
<tr>
<td>scenario</td>
<td>irony</td>
<td>upside</td>
</tr>
<tr>
<td>strategy</td>
<td>kicker</td>
<td>wager</td>
</tr>
<tr>
<td>chance</td>
<td>proof</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Less frequent nouns before is.

5.3 Expletive Subject Equivalents
The asterisk in Table 1 indicates that time was seems to work a bit differently than the other [Noun be] collocations. In addition to time being one of the few shell nouns that shows up only with the past tense verb, time was conveys a distinct meaning, as shown in the contrasting examples in (12).

(12) (a) [M]ore and more parents are after a name that stands out, that is actively, obviously different from other kids’ names. A name that’s cool. Time was, most kids were named what everyone else was named: for family members or saints, or following traditional ethnic or religious protocol. [COCA: Parenting, 2007]
(b) Something only angel warriors could do. Or so he’d thought. **Problem was,** she wasn’t an angel. [COCA: *The Darkest Secret*, 2011]

The difference arises from the intersection of aspect and givenness. The collocation *time was* sets up the existence of an era crucially having occurred in the past, as seen in the first line of each the paraphrases in (13). This highlighting of a referent existing and having been completed in the past cannot be similarly conveyed by *There was a problem/fact/question/truth*, as underscored by the adverb *once* added to these paraphrases.

(13) (a) *Time was,*
   "There (once) had been a time when"
   #$"There (now) was a time, which was"
   #$"It was a time when"

(b) *Problem was,*
   #$"There (once) had been a problem that"
   "There (now) was a problem, which was that"
   "It was a problem that"

*Fact was,*
   #$"there (once) had been a fact that"
   "There (now) was a fact which was that"
   "It was a fact that"

*Question was,*
   #$"there (once) had been a question whether"
   "There (now) was a question which was that"
   "It was a question whether"

*Truth was,*
   #$"there (once) had been a truth that"
   "There (now) was a truth, which was that"
   "It was a/the truth that"

Conversely, as seen in the second lines of the examples in (13), a way to capture the difference among [Noun was] tokens is that the present-day reading can be shown by the use of the adverb *now*, which works for all the nouns except *time*. This suggests a reading of the second-line examples as presentational, rather than existential constructions, despite the presence of the same copula verb in the first and second lines. (Without the adverbs, this reading is initially obscured because presentational forms are more often construed with non-*be* verbs, such as *showed up / appeared / stood / arose*).

Lastly, the distinctive characteristics of *time was* can also be captured in the ability to be recast as a sentence with an *it* expletive subject. To keep the same meaning as the pre-clausal *Time was* unit, *time* cannot be introduced as a new referent with an *it* sentence, since the time being discussed evokes an already existent, and completed era as its referent.
The other shell nouns, however, can be set up by a sentence that introduces the problem/fact/question, etc., with an indefinite noun form, as seen in the third row of examples.

These contrasts suggest that in the past tense form, the pre-clausal construction [Noun be] masks at least two underlying readings that can be distinguished by attempting to move the shell noun to a position after the copula.

6. Analysis of the Diachronic Corpus Data

Finally, following up on the idea that phrases becoming discourse markers might grammaticalize into a new function and position over time, I wanted to investigate the diachronic use of these pre-clausal units. Table 3 shows the results of my exploration of the occurrences in 25-year chunks between 1810 and 2012.

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Bare Shell Noun + is + punctuation</th>
<th>Bare Shell Noun + is + complementizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1810–1834 (COHA)</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>1835–1859 (COHA)</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>1860–1884 (COHA)</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>1885–1909 (COHA)</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td>1910–1934 (COHA)</td>
<td>57</td>
<td>8</td>
</tr>
<tr>
<td>1935–1959 (COHA)</td>
<td>67</td>
<td>18</td>
</tr>
<tr>
<td>1960–1989 (COHA)</td>
<td>108</td>
<td>15</td>
</tr>
<tr>
<td>1990–2012 (COCA)</td>
<td>1183</td>
<td>104</td>
</tr>
</tbody>
</table>

Table 3. Raw data for bare shell nouns.

However, there are fewer texts in each of the earlier eras of the corpus. So I have normed for number of hits per 1,000,000 in each of the rows. You can see the normed data in Table 4.

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Bare Shell Noun + is + punctuation</th>
<th>Bare Shell Noun + is + complementizer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1810–1834 (COHA)</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>1835–1859 (COHA)</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>1860–1884 (COHA)</td>
<td>0.54</td>
<td>0</td>
</tr>
<tr>
<td>1885–1909 (COHA)</td>
<td>0.72</td>
<td>0.018</td>
</tr>
<tr>
<td>1910–1934 (COHA)</td>
<td>0.94</td>
<td>0.13</td>
</tr>
<tr>
<td>1935–1959 (COHA)</td>
<td>1.10</td>
<td>0.296</td>
</tr>
<tr>
<td>1960–1989 (COHA)</td>
<td>1.46</td>
<td>0.20</td>
</tr>
<tr>
<td>1990–2012 (COCA)</td>
<td>2.48</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Table 4. Normed data for bare shell nouns.
In the 400 million words from 1810–1989, a total of 307 tokens of pre-clausal units with bare shell nouns were found, while in the 450 million words of COCA non-academic texts, 1183 tokens occurred. Even normed, a noticeable increase can be seen in the use of bare form, pre-clausal truth is type constrictions. The number of noun types involved also greatly increased. To see whether a relative change has occurred in bare forms compared to articulated forms, I also tracked the use of articulated shell nouns in pre-clausal units. These findings are shown in Table 5.

<table>
<thead>
<tr>
<th>Timeframe</th>
<th>Article + Shell Noun + is + punctuation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raw</td>
</tr>
<tr>
<td>1810–1834 (COHA)</td>
<td>310</td>
</tr>
<tr>
<td>1835–1859 (COHA)</td>
<td>833</td>
</tr>
<tr>
<td>1860–1884 (COHA)</td>
<td>811</td>
</tr>
<tr>
<td>1885–1909 (COHA)</td>
<td>579</td>
</tr>
<tr>
<td>1910–1934 (COHA)</td>
<td>481</td>
</tr>
<tr>
<td>1935–1959 (COHA)</td>
<td>371</td>
</tr>
<tr>
<td>1960–1989 (COHA)</td>
<td>642</td>
</tr>
<tr>
<td>1990–2012 (COCA)</td>
<td>8485</td>
</tr>
</tbody>
</table>

Table 5. Raw and normed data for articulated shell nouns.

Table 5 shows that the articulated form peaked earlier than the bare forms. The bare forms showed a slow but steady increase over the decades, while the articulated forms have slowly lessened in percentage, though the occurrence of both forms of the pre-clausal constructions has increased. This strengthens the argument that the pre-clausal forms are grammaticalizing into this new position and function, and away from a regular main clause use.

7. Conclusions

Future investigations of the bare forms in pre-clausal units will explore variation within the construction use, to determine which forms might be leading the change in function. Another direction to explore is whether different nouns are used more commonly in different registers and genres. Throughout, the approach will be to gather and contrast the behavior of the full range of nouns. For although Schmid (2000) covered many articulated shell nouns, most of the previous literature explores only individual shell noun types. For example, while noting that variants exist in which the “noun may range in semantic specificity from the minimal thing to semantically far richer nouns such as miracle,” Delahunty (2012, 43) examines only structures with the noun thing, Krug and Schützler (2013) look at idea, Aijmer (2007) looks at only the four nouns, fact, truth,
thing, trouble, etc. These individual case studies offer invaluable data points, which the present work builds on to show the benefit of examining the bare form of all the noun types found in this construction.

Through this initial overview of bare shell nouns in pre-clausal units, I have shown that they have limited referring abilities beyond identifying their complement clauses. Indeed, bare forms, especially those stranded from the verb in the unit, cannot be referred back to by pronouns. Distributionally, bare forms show up as subjects when they cannot show up after a verb, so referentially, these abstract bare forms function differently than bare concrete count nouns in discourse anaphora. In addition, looking across the full set of nouns used in the current pre-clausal form also allows us to unmask two underlying readings—an existential and a presentational construction—that can be distinguished by attempting to move the shell noun to a position after the copula. Finally, as a pre-clausal unit, the reduced forms, lacking both determiner and complementizer, are being used more often, and used more steadily across time. Together these findings support the idea that the truncated clauses are grammaticalizing.

**Works Cited**


**Corpora**
