Complex numerals in English: Constituents or not?[1]

CHUANSHENG HE
Hunan University

ONE-SOON HER
Tunghai University & National Chengchi University

This paper focuses on the controversy over whether complex numerals in English are constituents. Contra the traditional view (e.g., Hurford 1975; Greenberg 1990[1978]), the cascading structure proposed in Ionin & Matushansky (2006; 2018) maintains that cross-linguistically a complex numeral does not form a constituent to the exclusion of the NP complement. The derivation of an additive complex numeral, e.g., *twenty-two people*, thus involves an underlying source form with a nominal conjunction, e.g., *twenty people and two people*. Based on the argumentation established in He (2015) and He *et al.* (2017) supporting complex numerals as constituents in Chinese and minority languages in southern China, this paper first demonstrates that the non-constituency analysis is not viable for English, as the underlying forms of additive complex numerals may be ill-formed and also semantically inequivalent to the surface forms. We then offer evidence to support the constituency analysis from constituency tests and behavior of post-numeral and pre-numeral modifiers. Finally, we demonstrate that the extra mechanism of grafting, proposed by Meinunger (2015), is unnecessary for English complex numerals.

Keywords: cardinals, complex numerals, complementation, additive complex numerals

1 INTRODUCTION

According to standard assumptions concerning the syntax of numerals, numerals either occupy the specifier position of some functional head (e.g., Jackendoff 1977; Selkirk 1977; Franks 1994; Kayne 2005; Corver & Zwarts 2006; Stavrou & Terzi 2007, among others) or are analyzed as heads themselves (e.g. Ritter 1991; Giusti 1997; Rutkowski 2002; Borer 2005). Both assumptions are in line with the traditional view that a complex numeral, e.g., *two hundred* in *two hundred books*, functions exactly like a simple numeral, e.g., *two* in *two books*, and the grammatical relation between the multiplier in the complex numeral, i.e., *two* in *two hundred*, and the multiplicand, i.e., *hundred*, is that of modifier and head (e.g., Greenberg 1990[1978]: 287). Besides multiplication, a complex numeral may involve addition, e.g., *two hundred and twenty* in *two hundred and twenty books*, where *two hundred* and *twenty* form a numeral via conjunction. Thus, a complex numeral, like a simple numeral, is a constituent on its own, to the exclusion of the noun it modifies.

However, Ionin & Matushansky (IM hereafter) (2006) argue for a non-constituency treatment of complex numerals, based mainly on the overt case-marking phenomenon in Russian and other languages where numerals behave like heads assigning Case to the lexical NP, and also data from Turkish and other languages where lexical NPs with numerals larger than one require singular number-marking in spite of the availability of plural morphology. Ionin & Matushansky (2018) take this cascading structure as a basic tenet and further explore its implications cross-linguistically. These two opposing views are shown schematically in (1a) and (1b), respectively. The example used involves an additive complex numeral *two hundred and twenty*, which contains a multiplicative complex numeral *two hundred*.

(1) (a) Constituent account

two   hundred   and twenty   books

(b) Non-constituent account
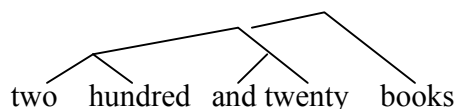
two   hundred   and   twenty books
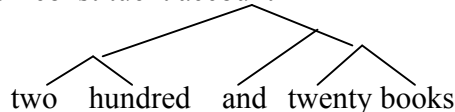
In IM's non-constituent account of (1b), in the surface form *two hundred and twenty* is not a constituent, *two hundred* is a constituent, and *twenty books* is a constituent. Yet, the coordinative structure [[two hundred] and [twenty books]] must somehow produce the semantics of 220 books. IM (2006) thus propose an underlying source form of [[two [hundred books]] and [twenty books]], as shown in (2b). The surface form in (2a) is derived from the source form either by right-node raising (RNR), as in (3a), or by PF deletion of the lexical NP in the first conjunct, as in (3a). IM (2018: 122, 136–7) further claim that, given that in some languages deletion is required for the derivation of additive numerals, posing right-node raising elsewhere is superfluous. For the sake of prudence, we shall still consider both options in deliberating the non-constituent account.

(2) (a) Surface form: [[two [hundred]] and [twenty books]]
    (b) Source form: [[two [hundred books]] and [twenty books]]

(3) (a) Right-node raising (RNR)
       [[two [hundred *t*]] and [twenty *t*]] <u>books</u>]

    (b) PF-deletion
       [[two [hundred ~~books~~]] and [twenty books]]

As shown in (2b) and (3a–b), the multiplicative complex numeral *two hundred* is also not a constituent in the source form. Thus, under the non-constituency account, a complex numeral, of any kind, is never a constituent to the exclusion of the NP. Note the crucial fact that an additive complex numeral may contain a multiplicative complex numeral, e.g., *two hundred* in *two hundred and twenty*, or more than one, e.g., *three thousand two hundred* is formed by conjoining the two multiplicative complex numerals *three thousand* and *two hundred*. Thus, logically if it can be demonstrated that an additive complex numeral like *two hundred and twenty* in (2a) is indeed a syntactic constituent to the exclusion of the NP *books*, then its two conjuncts, i.e., the multiplicative complex numeral *two hundred* and the simple numeral *twenty*, that it contains must also be constituents. Likewise, if multiplicative complex numerals are proven to be constituents, then additive complex numerals must also be constituents. This means that if a complex numeral, whether multiplicative or additive, can be proven to be a constituent in a language, then all complex numerals must be constituents in that language.

The non-constituency analysis of numerals is suggested to apply cross-linguistically. However, Meinunger (2015), using examples from English, German, Dutch, and Irish, presents syntactic, semantic, and pragmatic arguments against the non-constituency analysis, and proposes a graft (constituent) structure for complex numerals cross-linguistically. He (2015) and He *et al.* (2017) likewise argue convincingly that complex numerals in Chinese and a large number of minority languages in southern China are constituents. Her (2017) and Her & Tsai (2020) also demonstrate that in a numeral classifier construction the numeral, whether simple or complex, must be a constituent that merges with the classifier first. These studies thus cast doubts on the universality of the non-constituency analysis of complex numerals. After all, there are no reasons *a priori* to assume that complex numerals are universally constituents or non-constituents. A particular account justified for one or some languages cannot be automatically extended to all languages. In other words, whether syntactically complex numerals are constituents in a particular language is an empirical matter and should be determined based on linguistic facts in that language.

Noting that IM's non-constituency account is primarily motivated by data from languages other than English, we aim to demonstrate in this paper that, as far as English complex numerals are concerned, the non-constituent account is not viable. The paper is organized as follows. Section 2 first demonstrates that the non-constituent account runs into difficulty when we consider the grammaticality of the alleged source forms in various syntactic contexts. Section 3 then closely examines the issue of semantic (in)equivalence
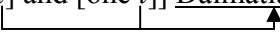
between the alleged source form and the surface form under the non-constituent account. In section 4, we provide more direct evidence in favor of the constituency analysis of complex numerals in English and demonstrate that grafting, proposed by Meinunger (2015), is unnecessary. Section 5 concludes the paper.

2 GRAMMATICALITY OF ALLEGED SOURCE FORMS

In this section we demonstrate that under the non-constituency view, the alleged source forms run into two kinds of problems. The problem with number agreement is discussed in 2.1, and 2.2 deals with a problem due to conjoined adjectives. In 2.3, we demonstrate that the non-constituency analysis also has difficulty accounting for post-numeral approximative markers such as *odd* and *or so*.

2.1 Problems due to number agreement

A general difficulty is found in additive complex numerals that end with the numeral *one*. The dilemma is while *one* requires a singular noun, the complex numeral requires the following NP to be plural (see Savu 2012 and Meinunger 2015: 109 for a similar observation). Consider example (4a) and its supposed underlying form in (4b). As shown in (5a–b), either RNR or PF-deletion is required to derive the surface form.

(4) (a) Surface: one hundred and one Dalmatians
    (b) Source: [one hundred Dalmatians] and [one Dalmatian]

(5) (a) [[[one hundred *t*] and [one *t*]] <u>Dalmatians</u>]

    (b) *[one hundred ~~Dalmatians~~] and [one Dalmatian]

    Consider RNR first. Though RNR is known to tolerate some form of plural mismatch, as shown in (6), the number morphology after movement must match with the final conjunct, not the first one. An RNR account predicts the ungrammaticality of *one hundred and one Dalmatians* and the well-formedness of *one hundred and one Dalmatian*, contrary to the fact.[2]

(6) (a) I have one, but you have one hundred books/*book.
    (b) I have one hundred, but you have one book/*books.

    Deletion fares no better, as the result is an ill-formed surface form (5b), *one hundred and one Dalmatian*. Thus, under both RNR and deletion, the desired source form of the second conjunct is in fact the ill-formed *one Dalmatians*. The problem is thus even more

---

[2] An anonymous reviewer suggests that the RNR option may possibly be adjusted to account for (5a), e.g., spell-out of the marked number feature among the number features of the two nominals or feature unification under conjunction (cf. The boy and the girl are/*is walking). Such measures are nonetheless stipulated technical solutions outside of the normal behavior of RNR.

obvious with plural nouns such as *people*. As shown in (7b), the alleged source form is ill-formed.

(7) (a) Surface: one hundred and one people
     (b) Source:*[one hundred people] and [one people]

2.2 Problems due to conjoined adjectives or nouns

The problem due to the numeral one in the second conjunct is made worse if the head noun is modified by a conjunction of two adjectives denoting attributes that are mutually exclusive. For example, (8a) denotes a group of neighbors, some friendly and others not, but the exact numbers of the two groups are unspecified. Consider its source form in (8b).

(8) (a) Surface: one hundred and one friendly and unfriendly neighbors
     (b) Source:*#[one hundred friendly and unfriendly neighbors] and
            [one friendly and unfriendly neighbors]

The problem of number agreement set aside, at least two neighbors are required in the second conjunct of (8b) for *friendly and unfriendly* to be felicitous. The PF deletion option for (8b) can thus be ruled out right away.[3] Under the RNR option, one might argue for an underlying structure of (9), where only the head noun is moved.

(9)
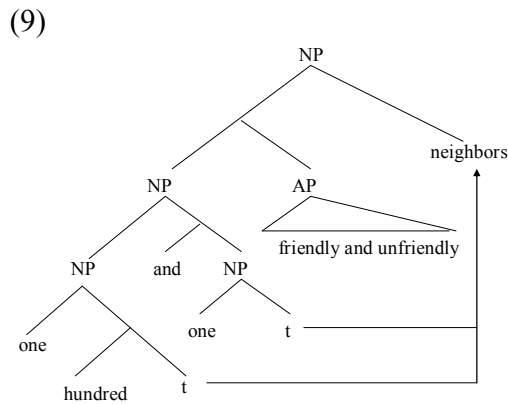
However, such an analysis derives a surface form where the adjectives and the complex numeral form a constituent, while the adjectives and the head noun do not. Such a left branching structure is surely not viable, as constituency tests can easily show that the adjective and the head noun form a constituent excluding the numeral, as in (10).

---

[3] An anonymous reviewer suggests that the problem may be removed if *neighbors* is not reconstructed at LF but is interpreted in the derived position. As we will argue more extensively in section 3, another problem of the source forms under the non-constituency account is that they are not semantically equivalent with the corresponding surface forms. The suggested measure may thus only be a stipulated technical solution to go around this specific problem due to complex numerals.

(10) (a) French and American workers, I have twenty-two.
       (b) I have twenty-two, not twenty-three, French and American workers.
       (c) I have twenty-two young French workers and old American workers.

Another undesirable consequence of (9) is that with stacked pre-nominal adjectives, we are led to a strange hierarchical structure for the adjectives. Though there has been much debate on the structural position of pre-nominal adjectives (as adjunct, head, specifier, or reduced relative clause) (e.g. Cinque 2005; 2010; Abels & Neeleman 2012; Talić 2017), the general agreement is that pre-nominal adjectives follow a hierarchy where the left one is structurally higher than its subsequent ones (for head-initial languages) (e.g. Vandelanotte 2002). Under the structure of (9), *twenty dangerous dead animals* is predicted to be [[[*twenty dangerous*] *dead*] *animals*], derived from [[[*twenty animals*] *dangerous*] *dead*]. However, *dead* and *animals* clearly form a constituent, as evidenced by *one*-replacement (i.e., *twenty dangerous ones* can mean *twenty dangerous dead animals*). The left branching structure of (9) makes incorrect predictions in scope interaction among different adjectives. It is usually the case that in a stacked adjectival structure, the left one takes a wider scope than the right one. For example, *a dead dangerous animal* might be a dead lion, which was dangerous in life but is now dead, while *a dangerous dead animal* might be a dead sheep which is infected with dangerous viruses (see Svenonius 1994 for more examples and discussions).

Note also that the problem is not restricted to only numerals ending in *one*. In general, such a problem arises when the ending numeral in an additive denotes a number that is smaller than the number of distinguishing adjectives modifying the noun. For example, the second conjunct in (11b) is uninterpretable, where the number is *two* and yet three distinguishing adjectives are used.

(11) (a) Surface: one hundred and two Chinese, Korean, and Japanese neighbors
       (b) Source:*#[one hundred Chinese, Korean, and Japanese neighbors] and
              [two Chinese, Korean, and Japanese neighbors]

This kind of mismatch thus also exists in examples where the head noun consists of two or more conjuncts. Consider (12a) and note the second conjunct in its alleged source form (12b), i.e., *one men and women*, which is uninterpretable, for obvious reasons. The problem is again not limited to numerals ending in *one*. Consider (13) and (14).

(12) (a) Surface: one hundred and one men and women
       (b) Source:*#[one hundred men and women] and [one men and women]

(13) (a) Surface: one hundred and two men, women, and children
       (b) Source:*#[one hundred men, women, and children] and
              [two men, women, and children]

(14) (a) Surface: one hundred and three men, women, children, and infants
       (b) Source:*#[one hundred men, women, children, and infants and
              [three men, women, children, and infants]

If the numeral ends in *two*, the same mismatch appears when the head noun is an element having three or more conjuncts. In principle, this kind of mismatch exists in English with numerals ending in *n*, $1 \leq n < 10$, while the head noun consists of $n+1$ or more conjuncts.

2.3 Problems due to post-numeral approximatives

Numerical expressions can be modified, e.g., *exactly sixty seconds* or *around twenty-four hours* (e.g., Plank 2004). A small class of such modifiers in English appear immediately after a numeral, including clitics such as *-ish*, *-odd*, and *-plus*, and phrases such as *or so*, *or more*, and *or thereabouts*, to express numerical approximation. We will use *-odd* and *or so* as examples and leave aside the rest. Note that *-odd* can appear at the end of a numeral and before the noun, as in (15a), or between two numeral bases within a complex numeral, as in (15b).

(15) (a) one hundred million odd people
      (b) one hundred odd million people

Hamawand (2017: 108) treats *-ish* and *-odd* as suffixes; however, a closer examination shows that the host they are attached to must be a phrase, not a morpheme. For example, in *seven o'clock-ish* the bound lexical form *-ish* is attached to the phrase *seven o'clock*, not to *o'clock*, hence the same interpretation as *approximately seven o'clock*. Likewise, *one hundred odd people* means *slightly over one hundred people* (Hamawand 2017: 109). Thus, the range of 'slightly over' is interpreted in relation to the entire number, not the single word *-odd* is attached to. To illustrate, *one hundred million-odd people* can be one hundred million people plus one percent, thus plus one million people, but *one million-odd people* cannot be one million people plus another one million people.

A straightforward account is thus to take *odd* as an adverbial modifying the numeral *one hundred million* as a constituent in (15a) and *one hundred* as a constituent in (15b), the same way Pankau (2018) accounts for approximative numerals in German. It is not clear how a non-constituency account would deal with this, as IM (2006; 2018) have not discussed post-numeral approximatives. Consider (15a) first. Assuming a coordinative structure between the numeral and *odd*, possible candidates of its appropriate underlying form are listed in (16). Notice that in all such forms, the second conjunct is ill-formed and uninterpretable.

(16) (a) *one hundred million ~~people~~ (and) odd people
      (b) *one hundred million ~~people~~ (and) odd ~~million~~ people
      (c) *one hundred million ~~people~~ (and) odd ~~hundred million~~ people
      (d) *one hundred million ~~people~~ (and) odd ~~one hundred million~~ people

Note also that, as Kayne (2012: 78) observes, in a [numeral NP] phrase in English, the NP can readily be deleted, but not the numeral, e.g., *Mary has written four papers, whereas*

John has only written four squibs, where the latter *four* must be pronounced. This means that (16b-d), where a numeral is deleted, are all ill-formed.

Now consider (15b). Under the non-constituency analysis, three candidate source forms are possible, as in (17), but all are ill-formed and without a congruent meaning.

(17) (a) *one hundred ~~million people~~ (and) odd million people
    (b) *one hundred ~~million people~~ (and) odd ~~hundred~~ million people
    (c) *one hundred ~~million people~~ (and) odd ~~one hundred~~ million people

Like *odd*, *or so* can appear at the end of a complex numeral, as in (18a), or within a complex numeral, as in (18b). There are two possible analyses of *or so*, i.e., lexical or phrasal. If *or so* is a lexical item like *-odd*, then the same argument applies. However, we agree with Hamawand (2017: 110) that *or so*, *or more*, and *or thereabouts* are (idiomatic) phrases instead and *or* here is a conjunction. Then the possible source forms of (18a) and (18b) under the non-constituency account are those in (19) and (20), respectively.

(18) (a) one hundred million or so people
    (b) one hundred or so million people

(19) (a) *one hundred million ~~people~~ or so people
    (b) *one hundred million ~~people~~ or so ~~million~~ people
    (c) *one hundred million ~~people~~ or so ~~hundred million~~ people
    (d) *one hundred million ~~people~~ or so ~~one hundred million~~ people

(20) (a) *one hundred ~~million people~~ or so million people
    (b) *one hundred ~~million people~~ or so ~~hundred~~ million people
    (c) *one hundred ~~million people~~ or so ~~one hundred~~ million people

Clearly, none of the candidate source forms in (19) and (20) is viable, given that they all are ill-formed syntactically. The facts discussed thus far all cast doubts on the non-constituency analysis of English numerals. In the next section, we will provide more evidence from semantics.

3   SEMANTIC (IN)EQUIVALENCE BETWEEN SOURCE AND SURFACE

To support their proposal that additive numerals are derived from coordinated NPs, IM (2006: 342) cite examples from Luvale (Zweig [2006]) and Biblical Hebrew, where the lexical NP appears in both conjuncts of an additive numeral.

(21) mikoko makumi atanu   na-mikoko   vatanu
     sheep   ten   five   and-sheep   five
     'fifty-five sheep' (Luvale, IM's (2006) (46))

(22) tēšaʕ šanîm   u-   mātayim   šānā. . .
     nine year-PL   and   hundred-DU   year

'<And Peleg lived after he begat Reu> two hundred and nine years' (Genesis, 11.19, IM's (2006) (49))

The data in (21) and (22) suggest that *fifty-five sheep* and *two hundred and nine years* in English can likewise be analyzed as *fifty sheep and five sheep* and *two hundred years and nine years,* respectively. This is not a strong argument, however, considering the fact that such forms co-exist with the more canonical forms with only one occurrence of the lexical NP and the former may be used for special pragmatic effects. More importantly, for this argument to work, the truth-conditional equivalence must be established between the surface form and the proposed source form. As clearly established by Her & Tsai (2014; 2015), whether the derivation of the surface from the source is by movement or by ellipsis, the two forms must be semantically equivalent. Consider the examples in (23). In the non-constituency account, (23a) has the underlying form in (23b).

(23) (a) Surface: Twenty-two people came to the party.
      (b) Source: Twenty people and two people came to the party.

IM (2018: 330) consider (23b) acceptable under the interpretation of two distinct entities. IM (2006: 347) suggest that the subtle difference between (23a) and (23b) may be accounted for under the maxim of Manner, which prefers the more succinct (23a) over the lengthier (23b), but such a conversational maxim can be overridden by the pragmatic need to separate the two groups. Crucially, IM (2006: 347) explicitly claim that the pair 'have the same truth-conditions'.[4] Nonetheless, the apparent semantic equivalence is merely accidental. Consider the pair in (24).

(24) (a) Surface form: Twenty-two students each have a different book.
      (b) Source form: [Twenty students] and [two students] each have a different book.

By the internal reading of *different* (e.g., Carlson 1987), (24a) indicates that there are 22 different books and each student has one, while (24b), to the extent that it is acceptable, may indicate that there are only 2 different books, of which 20 students have the same one and 2 students have a different one. Therefore, a scenario exists where (24b) is true and (24a) is false.

We will provide a number of more cases that challenge the semantic equivalence between the surface forms of additive numerals and their alleged source forms.

3.1 Problems due to conjoined adjectives or nouns

---

[4] As pointed out by an anonymous reviewer, (23a) has an additional reading of event counting, in the sense of Krifka (1990). That is, *22 people came to the party* can mean that there were 22 visits, thus by 22 people or less, as some may have come and left and come back. In this reading 22 does not quantify over individuals, but over events. IM do not take this reading into account, and nor will we, as this reading is not crucial for the debate.

Consider examples involving coordinated numerals and coordinated head nouns. (25b) is the supposed source form of (25a), and the two possible derivations, RNR and PF-deletion, are shown in (26a) and (26b), respectively.

(25) (a) Surface form: one hundred twenty men and women
      (b) Source form: [one hundred men and women] (and) [twenty men and women]

(26) (a) [[[one hundred *t*] and [twenty *t*]] men and women]

      (b) [one hundred ~~men and women~~] and [twenty men and women]

According to Heycock & Zamparelli (2005: 246), it is possible for *one hundred twenty men and women* in (25a) to refer to a group of 119 men and 1 woman or a group of 119 women and 1 man.[5] But these two readings are not available in (25b), where the denotation of *one hundred men and women* should contain at least 1 man and 1 woman, and the denotation of *twenty men and women* should likewise contain at least 1 man and 1 woman. Thus, there must be at least 2 men and 2 women in every possible reading.

The same problem is found in example (27a) and its source form (27b). Again, the two extreme readings of (27a) are a group of 119 friendly neighbors and one unfriendly neighbor or a group of 119 unfriendly neighbors and one friendly neighbor. These two readings are missing in (27b), where the modified head noun *friendly and unfriendly neighbors* is base-generated in two conjuncts, which requires that there are at least two friendly neighbors and at least two unfriendly neighbors in every interpretation.

(27) (a) Surface form: one hundred twenty friendly and unfriendly neighbors
      (b) Source form: [one hundred friendly and unfriendly neighbors] (and)
                       [twenty friendly and unfriendly neighbors]

Thus, we may conclude that in terms of truth condition (denotations, more precisely), it is a mistake to assume that additive numerals are derived from NP coordination with occurrences of the lexical NP in each conjunct. In general, the denotational equivalence between additive numerals and their alleged sources cannot be established. The reading possibilities of the former in general outnumber those of the latter.

3.2 Problems due to reciprocal and distributive expressions

We now consider complex numerals with reciprocals, e.g., *each other* and *one another*, and distributive expressions, e.g., *each* and *respectively*. The overall potential problem is the meanings allowed by the source forms are not the same as those of the surface forms.

---

[5] This kind of extreme reading is easily available in Chinese, which does not have plural morphology. The following Chinese counterpart (i) can denote either a group of 119 men and 1 woman or a group of 119 women and one man.
(i) yi-bai   er-shi   ge   nanren   he   nüren
    one-hundred twenty   Cl   man   and   woman
    'one hundred twenty men and women'

Again, as IM (2006; 2018) have not discussed these items, we can only assume the source forms they have proposed elsewhere.[6]

The first item we shall consider is the reciprocal expression *each other*. The surface form (28a) favors strong reciprocity for *each other* (e.g., Langendoen 1978). If two students' liking each other is counted as one pair of mutual liking, then, given that each of the twenty-two students likes each other, there are 231 pairs (0+1+2+3+…+21).

(28) (a) Surface form: Twenty-two students like each other.
    (b) Source form: [Twenty students] and [two students] like each other.

However, the source form (28b) does not mean the same in that it involves only 191 pairs of mutual liking ([0+1+2+3+…+19] + [1]). In addition, (28b) allows the reading that the first group of 20 students and the second group of two students like each other, but students within a group may or may not like each other. Such a reading is impossible in (28a). Now, consider expressions with *each*; see the pair in (29).

(29) (a) Surface: Twenty-two people came to the two parties each.
    (b) Source: Twenty people and two people came to the two parties each.

In (29a), the total number of people is 44, as each party has 22 people. Yet, the source form (29b) has an additional reading that one party has 20 people and the other has 2. The problem gets worse when the number of corresponding pairs is larger than two, as shown in (30).

(30) (a) Surface: Two hundred twenty-two people came to the three parties each.
    (b) Source: [Two hundred people] and [twenty people] and [two people]
              came to the three parties each.

The same problem exists with the adjective *respective*. In the surface (31a), the total number of students is at least 44 (22×2), but the source (31b) allows a reading where the number is 22 (20+2).

(31) (a) Surface form: The two teachers love their respective twenty-two students.
    (b) Source form: The two teachers love their respective [twenty students] and
              [two students].

The problem described here may arise even in contexts without overt *respective* or *respectively*. Consider (32).

(32) (a) Surface form: From the two jobs, he earned twenty-two dollars.
    (b) Source form: From the two jobs, he earned twenty dollars and two dollars.

---

[6] An anonymous reviewer states that perhaps IM's conjunctions are not supposed to correspond to the natural-language conjunctions in the putative source examples, but we have no way of knowing.

The reading of (32b) is that, between the two jobs, one job paid $20 and the other $2, but such a reading is merely one of the numerous possibilities in (32a), where the only thing for sure is that the total amount paid for the two jobs is $22. The examples in (33) present another case.

(33) (a) Surface form: He is paid every twenty-eight days.
      (b) Source form: He is paid every [twenty days] and [eight days].

Note that (33b) has the reading that he is paid every 20 days and every 8 days, but this reading is not available in (33a). The two forms thus have different meanings. The final case we present is (34).

(34) (a) Surface form: The price is between ten and thirty-three dollars.
      (b) Source form: The price is between [ten] and [thirty dollars] and [three dollars]

There is only one reading in (34a): $10-$33. However, (33b) is ambiguous, as there are two instances of *and*, and each can correspond to *between*. The price can thus be either between $10 and $33 (30+3) or between $40 (10+30) and $3.

3.3 Problems due to post-numeral approximatives

Note again that IM (2006; 2018) have not discussed post-numeral approximatives such as *-ish*, *-odd*, *-plus*, *or so*, *or more*, and *or thereabouts*, and it is thus unclear how a non-constituency account would deal with them. In 2.3 we have demonstrated that under the non-constituency analysis, it may be difficult to come up with a grammatically well-formed source form for numerical expressions with such an approximative marker. We will now demonstrate, with *or so* as an example, that it is likewise difficult to come up with a source form that is semantically equivalent to the surface form. Recall that *or so* can appear between the numeral and the noun or between two numeral bases within a complex numeral, as shown in (35a) and (36b), respectively.

(35) (a) one hundred or so people
      (b) one hundred or so million people

Given the cascading structure of [one [[hundred] people]]], *or so* in (35a) can form a constituent either with *hundred*, as in [one [[hundred or so] people]]], or with *people*, [one [[hundred] [or so people]]]]. Both options are undesirable. Hypothetically, one may propose to derive *100 or so people* from the well-formed form *100 people or so*, and *100 or so million people* from the well-formed *100 million people or so*, by rightward movement of the NP, as in (36).

(36) (a) one hundred ~~people~~ or so people
                        └_____↑

      (b) one hundred ~~million people~~ or so million people
                        └_____↑

However, such movements are clearly too powerful and also violate the coordinate structure constraint. One may argue that *or so* is not phrasal and *or* here is not a conjunction. In other words, *or so* may be seen as a single lexical item modifying *hundred*; [*hundred or so*] is thus a constituent. The two examples in (35) thus have the structure shown in (37).

(37) (a) [one [[hundred or so] people]]
       (b) [one [[hundred or so] [million people]]]

However, this analysis leads to dubious interpretations of a [numeral *or so*] expression. As noted in Corver & Zwarts (2006: 831), the numeral in combination with the sequence *or so* obtains an approximative interpretation: 'twenty or in the vicinity of twenty' for *twenty or so*. Let's say that *or so* indicates a portion in the range of minus and plus 10% of the preceding numeral base. If the numeral base is *hundred*, *hundred or so* would denote 90-110 (100±10). If the numeral base is *thousand*, *thousand or so* would denote 900-1100 (1000±100). Given this interpretation of *or so*, the equations in (38) hold according to the semantics proposed in IM (2006), e.g., $3 \times (90\text{-}110)=(270\text{-}330)$, $9 \times (90\text{-}110)=(810\text{-}990)$, etc. The argument applies equally to *one hundred or so million people*.

(38) (a) one [[hundred or so] people]=90-110 people
       (b) two [[hundred or so] people] =180-220 people
       (c) three [[hundred or so] people] =270-330 people
       (d) ……
       (e) nine [[hundred or so] people] =810-990 people

However, our informant indicates that the correct interpretations should be the ones shown in (39), where *or so* would denote a range of minus and plus 10% of the numeral base *hundred*, irrespective of the preceding numerals (*one*, *two*, *three*, … *nine*). That is, *two hundred or so people* denotes 200±10=190-210 people and *nine hundred or so people* denotes 900±10=890-910 people.

(39) (a) one hundred or so people =90-110 people
       (b) two hundred or so people =190-210 people
       (c) three hundred or so people =290-310 people
       (d) ……
       (e) nine hundred or so people =890-910 people

This intuition is easy to verify. Consider (39e), *nine hundred or so people* should not mean a range of 810-990 people, as in (38e). If the number is 810, it is unlikely that we say *nine hundred or so*. Rather we would say *eight hundred or so* (eight hundred or so people = 790-810 people). If the number is 990, it is also unlikely that we say *nine hundred or so*. Instead we would say *one thousand or so*.

The discussion above indicates that it is difficult to come up with a source form that is semantically equivalent to the surface form with *or so*.

3.4 Problems due to bare numerals and the silent noun

A crucial consequence of the non-constituency analysis is that a bare numeral phrase is never bare, as there is a silent noun as complement. Every numeral, simple or complex, thus must be a nominal projection. This means that numbers per se cannot be referred to directly; rather it is the number of *things* that is referred to instead. We will demonstrate with different kinds of examples that this silent *things* with bare numerals leads to various kinds of semantic oddities. We will begin by considering the two mathematical expressions involving addition in (40).

(40) (a) Two and two are four.
    (b) Two and two is four.

Partially following Hofweber (2005), IM (2006: 353) suggest that both expressions involve a silent X, as in (41).

(41) (a) For whatever X, two X and two X are four X.
    (b) For whatever X, two X and two X is four X.

The paraphrases in (41) seem reasonable. This can be shown by the example with continuation in (42), where X is replaced by a lexical noun.

(42) For whatever X, two X and two X are/is four X. If X is a car, then two cars
    and two cars are/is four cars.

Nevertheless, this treatment cannot be extended beyond addition and subtraction. Consider division and multiplication in (43), where the two verbs *divide* and *multiply* select numbers, not things, as arguments. The paraphrases in (44) thus make no sense.

(43) (a) If you multiply ten by five, you get fifty.
    (b) If you divide ten by five, you get two.

(44) (a) For whatever X, if you multiply ten X by five X, you get fifty X.
        ??If X is a car, then if you multiply ten cars by five cars, you get fifty cars.
    (b) For whatever X, if you divide ten X by five X, you get fifty X.
        ??If X is a car, then if you divide ten cars by five cars, you get two cars.

Furthermore, this treatment fails to cover other arithmetic contexts. One context concerns *the number of planets*. Consider the example in (45), which IM (2006: 353) also mention but no discussion is provided on how to handle it.

(45) The number of planets within the Solar System is nine.

Under the non-constituent account, the numeral in (45) is a nominal projection with a silent noun X, as shown in (46). The problem is that, unlike the paraphrases (41), (46) is uninterpretable, because *nine X (for whatever X)* is semantically incompatible with the

subject *the number of planets*, as shown in (47), supposing that the silent noun is the general word THINGS. This is because what *nine X* denotes belongs to the domain of things. We cannot say that the number of planets is nine things, whatever things they may be. Note that, following the convention in the literature, capital letters indicate silence; THINGS in (47) is thus the silent counterpart of *things* (e.g., Her & Tsai 2015).

(46) The number of planets within the Solar System is nine X, for whatever X.
(47) #The number of planets within the Solar System is nine THINGS,
      for whatever THINGS.

Unlike Hofweber (2005), who proposes a cognitive type-coercion operation that may convert the quantifier phrase *nine* (hence a constituent alone) into a singular term,[7] in IM (2006), such type-lowering operation or Partee's (1986) type-shifting operation does not help, because the numeral *nine* should be derived from *nine X* (or in case of complex numerals, say *one hundred and nine*, it should be derived from [[[*one hundred t*] and [*nine t*]] *X*] or [[*one hundred X̶*] *and* [*nine X*]]). Type-shifting cannot change the fact that *nine X* cannot refer to a number.

Similarly, the following arithmetic sentence encounters the same semantic oddity if numeral phrases are nominal ones in disguise.

(48) (a) The square root of one hundred is ten.
      (b) #The square root of one hundred THINGS is ten THINGS.

Another context concerns explicit number-referring terms such as *the number eight* or *the number ninety-nine*. According to Moltmann (2011), explicit number-referring terms are genuine singular terms taking direct reference to numbers, by means of which we are able to talk about numbers. It makes little sense to interpret *the number eight* as *the number eight X* or *the number ninety-nine* as *the number ninety X* (*and*) *nine X*.

Similarly, in the following mathematical expression (49), we cannot insert any silent noun behind the numeral, which might result in abnormality in semantic selection, for THINGS is *not* a number (see Hurford 1987: 159 for similar examples and discussions).

(49) (a) Seven is a prime number.
      (b) #Seven THINGS is a prime number.

Following the logic that the term *number* must refer to the numeral, simple or complex, as a separate unity without the noun, we contend that the formation of an ordinal number in English must be based on a corresponding cardinal number as a constituent on its own. IM (2006:350; 2018:7) have specifically left aside ordinals, but as an anonymous reviewer points out, when it comes to ordinal numbers, it is very hard, if not impossible, to argue for

---

[7] Hofweber's account of sentence (45) is not by a type-lowering operation which is proposed to handle sentences like *two and two is four*. According to Hofweber (2005), (45) is supposed to be derived from some kind of focus construction. In Moltmann (2011), sentence (45) is treated as a specificational sentence (Higgins 1973), with *the number of*-terms as concealed questions and *nine* as expressing a plural property. Neither account has the bare numeral require a silent noun.

a non-constituency approach. Consider these two examples, *the year two thousand and twenty-one* and *the two thousand and twenty-first year*. It is indeed very hard, if not impossible, to argue for such source forms: *the year two thousand* YEAR *and twenty* YEAR & *one* and *the two thousand* YEAR *and twenty* YEAR & *first year*. In a constituency approach, such problems disappear.

Finally, also suggested by an anonymous reviewer, the stress pattern of a bare numeral is different from that of an NP with a numeral modifier. In the phrase *twenty-two books* for example, the main stress falls on the noun, indicated with uppercase letters, and a side accent on the initial number: <twénty seven BOOKS>. However, in the case of the bare numeral, the main accent goes to the final digit, the initial potency being unstressed: <twenty TWO>.

Another anonymous reviewer points out that the arguments put forth in this subsection cannot serve as arguments against the non-constituency analysis of the [numeral noun] construction, only against the extension of this analysis to bare numerals in mathematical statements. We agree. However, recall that IM's extension of [numeral noun] structure to bare numerals is motivated by the claim that complex numerals are never constituents. The arguments in this subsection clearly show that bare numerals are constituents.

The serious challenges to the non-constituency analysis of English numerals suggest that it is necessary to return to the traditional view that treats complex numerals as constituents, whether there is an ensuing lexical NP or not.

4   COMPLEX NUMERALS AS CONSTITUENTS

We first present evidence from constituency tests in section 4.1, and then in 4.2 and 4.3 we offer evidence from post-numeral and pre-numeral approximative markers. In 4.4, we briefly discuss the straightforward multiplicative complex numerals, e.g., *two hundred*, without addends, and argue for their constituency. In 4.5, we demonstrate that complex numerals in English are syntactic constituents straightforwardly and thus the grafting account proposed by Meinunger (2015) is not justified for English.

4.1 Evidence from constituency tests

There is evidence from conventional constituency tests indicating that numerals are constituents. Consider substitution first.

(50) (a) A: They say the king has *many* children.
       (b) B: Yes, he has *twenty-two* children.

(51) (a) A: Tell me *how many* children the king has.
       (b) B: He has *twenty-two* children.

(52) (a) A: Do you know *how many* thousand soldiers the king has?
       (b) B: Yes, *twenty-two*.

In the examples in (50)–(52), the complex numeral *twenty-two* can be substituted, and referred to, by the quantifier *many* (50a) and the interrogative *how many* (51a, 52a). In

addition, (52b) also provides evidence from the fact that *twenty-two* stands alone as a short answer. (52a) is especially telling, where the multiplier alone of a multiplicative complex numeral can be the target of *how many*.

Now compare (53a) and (53b), where the correspondence between the pair is similar to that of pseudo-clefting, again showing that *twenty-two* is a constituent. In (54), we see that two numerals, complex or not, can be coordinated, which is also consistent with their syntactic status as constituents.

(53) (a) The king has twenty-two children.
    (b) Twenty-two is the number of children the king has.
    (c) The number of children the king has is twenty-two.

(54) (a) It makes a big difference whether he has two or twenty-two children.
    (b) It makes no difference whether he has twenty-one or twenty-two children.

Another piece of evidence comes from compounds that involve a numeral and a lexical noun. In (55) and (56), the noun in question enters the [numeral + noun(-*ed*)] word formation process as a bare lexical item, and thus does not carry plural morphology in spite of the plurality denoted by the numeral. The numeral entering the word formation process thus must also be either a lexical item, in the case of simple numerals, e.g., *three* and *seven*, or, a phrasal constituent in the case of complex numerals, e.g., *three hundred* and *thirty-seven thousand*, whose internal syntactic structure is widely recognized.

(55) (a) a one-hundred-dollar bill
    (b) a twenty-four-bone umbrella
    (c) a thirty-two-foot sailboat
    (d) a four-hundred-and-four-carat diamond
    (e) a twenty-six-letter word

(56) (a) a three-hundred-legged centipede
    (b) a two-hundred-headed monster
    (c) a three-footed stand
    (d) a two-handed sword
    (e) a five-armed starfish

4.2 Evidence from post-numeral approximatives

In 3.3 we have demonstrated that post-numeral approximative markers such as *plus*, *odd*, -*ish*, and *or so* must be interpreted in terms of the numerical value denoted by the entire numeral that precedes it, not the preceding numeral base. This generalization can be further confirmed by the fact that, though the prototypical use of such approximative markers is to co-occur with round numbers, i.e., numbers that end with a numeral base, in certain domains it is perfectly natural for such markers to appear with non-round numbers (Dooley & de Haan 2008: 39), as shown in the three sets of examples in (57)–(60), obtained from Google search. The emphasis by italics is added by the authors.

(57) (a) She glanced at the clock, *twenty-plus* hours to go.
      (b) The *twenty-two plus* hours from Dallas to Denver was absolutely awful.

(58) (a) *Twenty-odd* hours in 120-degree heat can play a few tricks on the mind.
      (b) *Twenty-two odd* hours later, with neither of us resting, the work was done.

(59) (a) We have basically *twenty-ish* hours.
      (b) *Twenty-two-ish* hours to go 'til the poll, for anyone else who wants to enter.

(60) (a) *Twenty or so* hours later, we arrived back in good old New Bedford.
      (b) Not once during the *twenty-two or so* hours I was there did the hospital offer
          me food or water.

While *plus* and *odd* express a narrow range above, not below, the cardinality of the modified numeral, *-ish* and *or so* indicate that the actual number may be either slightly above or below the reference number (Dooley & de Haan 2008: 37). The crucial point that the examples in (57)–(60) demonstrate is that the actual range of cardinality such an approximative marker expresses is interpreted in terms of the numerical value denoted by the entire (complex) numeral that precedes it. The (complex) numeral that precedes such markers thus must form a syntactic constituent.

4.3 Evidence from the modified cardinal construction

In the case of modified numerals, some modifiers occur before the numeral, e.g., *more than ten books*, *around ten books*, *exactly ten books*, etc. Unlike the standard analysis in which the modifier and numeral form a constituent to the exclusion of the lexical NP, as shown in (61) (Barwise & Cooper 1981; Keenan & Stavi 1986; Kadmon 1987; Hackl 2000), the syntax that IM (2006; 2018) propose requires that the numeral combine with the lexical NP before combining with the modifier, as in (62).

(61) [[over ten] books]

       over   ten   books
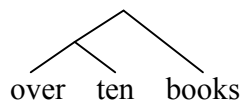
(62) [over [ten books]]
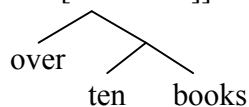
       over
              ten   books

The syntax of such modified numerals has been controversial. Some works have argued that in a numeral-containing comparative (*more than ten books*), the numeral first combines with the lexical NP (Arregi 2010, among others). Yet, Corver & Zwarts (2006) argue that for numeral constructions such as *over ten books*, *over* and the numeral *ten* form

a constituent excluding the lexical NP. On the other hand, IM (2006: 352) justify why (62) is to be preferred over (61) with two arguments. First, consider *just about ten books*. For this phrase, which they consider a PP, P must be linearized before the numeral, which takes the noun as complement. Second, *ten books* can be replaced by an NP such as *the predicted number of books*, as in (63), thus supporting the view in (62), that *ten books* is a constituent.

(63) [over [the predicted number of books]]

However, this second argument of IM's does not negate the standard assumption that modifiers can directly modify a numeral, because the phrase *the predicted number* can denote a number, e.g., *ten* or *twenty-two*, as in *the predicted number is ten/twenty-two*. Further, IM's first argument, if we understand it correctly, is that [*just about ten books*] is a PP. Yet, as Corver & Zwarts (2006) argue, for expressions like *over ten books*, *over ten* is best seen as having the same distribution and function as a bare numeral. Though their focus is on Dutch, their syntactic and semantic arguments apply to English as well. Particularly, expressions like *over ten books* behave like noun phrases instead of prepositional or adverbial phrases. This can be confirmed by the fact that they occur in various syntactic positions that require NPs, as shown in (64).

(64) (a) His over ten books were all bestsellers.
      (b) Over ten books are needed for the class.
      (c) He has written over ten books.
      (d) He has given me over ten books.
      (e) The same character has appeared in over ten books.

As demonstrated above, *over ten books* is the possessed NP in (64a), the subject in (64b), the direct object of a transitive verb in (64c), the secondary object of a ditransitive verb in (64d), and the object of a preposition in (64e). These facts support the view that modified numeral expressions such as *over ten books*, *around ten books*, *exactly ten books*, etc. should be seen as NPs. Under the PP analysis of *over ten books*, the sequence *over ten* cannot be a constituent. Yet, under the NP analysis where *over* can be seen as an adverbial, much like *approximately*, modifying the numeral *ten*, the sequence *over ten* thus forms a constituent excluding the lexical NP. As a constituent, *over ten* can be replaced straightforwardly by either a simple numeral or a complex numeral, as in (65).[8]

(65) (a) His eleven/twenty-two books were all bestsellers.
      (b) Eleven/twenty-two books are needed for the class.
      (c) He has written eleven/twenty-two books.
      (d) He has given me eleven/twenty-two books.
      (e) The same character has appeared in eleven/twenty-two books.

---

[8] Replacement, e.g., *one*-replacement for NP and *do-so*-replacement for VP, is a well-accepted constituency test, under the condition that the meaning is kept as closely related to the original as possible (Carnie 2013: 98). Importantly, replacement can only be seen as a necessary condition of constituency, not a sufficient condition.

However, the most important evidence in favor of the constituent account is from comparative numerals, e.g., *more than five sandwiches*. IM (2018: 300) aptly point out, such modified numerals may have two different readings, as in (66), which can be brought out in appropriate discourse contexts, as in (67).

(66)   more than five sandwiches
   (a) 'Many' reading ≈ 'six or more sandwiches'
    (b) 'Much' reading ≈ 'five sandwiches plus something else'

(67)   (a) 'Many' reading: I ate more than five sandwiches. I ate six!
   (b) 'Much' reading: I ate more than five sandwiches. I also ate a bowl of soup!

Though IM (2018: 302) do recognize that the two readings may be the result of a structural ambiguity, i.e., the 'many' reading involves 'more than five' as a constituent, while the 'much' reading is due to 'five sandwiches' as a constituent, as in (68a–b), they argue that (68a) is unnecessary and is thus rejected, for (68b) alone already allows for both readings, as the 'many' reading can be subsumed in the 'much' reading, i.e., 'five sandwiches plus something else' could be six sandwiches.

(68)   (a) 'Many' reading: [[more than five] sandwiches]

        more than   five   sandwiches

    (b) 'Much' reading: [more than [five sandwiches]]
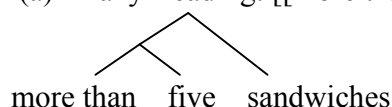
    more than
            five   sandwiches
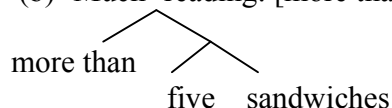
However, a seemingly trivial but critical fact has been overlooked in previous studies, i.e., in both (68a) and (68b) the simplex numeral five is in fact a constituent, which can be replaced with a complex numeral. The two structures are exactly the ones allowed under the constituent account. Consider (69a–b), where the simplex numeral *five* as a constituent has now replaced a complex numeral *five hundred* as a constituent; the two structures thus also account for the 'many' and 'much' readings, respectively. IM's cascading structure of (70), where the complex numeral is not a constituent, can accommodate the 'much' reading only, structurally. That (70) is ambiguous with two readings, with the 'many' reading subsumed in the 'much' reading, is thus just a claim that has not been unsubstantiated.
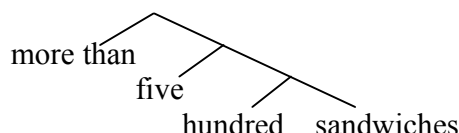
(69)   Constituent account
    (a)

    more than

```
            five    hundred    sandwiches
      (b)
     more than
              five   hundred    sandwiches
```

(70) Non-constituent account

```
           more than
                 five
                     hundred    sandwiches
```
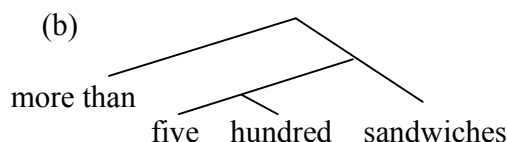
We thus argue for the opposite, that it is the non-constituent account of (70) that is unnecessary and should thus be rejected. The 'many' and 'much' readings of comparative numerals in English can be adequately and transparently accounted for under the constituent account via a structural ambiguity, as in (69a–b).[9]

4.4 More on multiplicative complex numerals

Some of the arguments for complex numerals as constituents only use additive complex numerals such as *two hundred and twenty books* as examples. Note that complex additive numerals like *two hundred and twenty* also contain a multiplicative complex numeral *two hundred*. Thus, crucially, if the additive complex numeral *two hundred and twenty* is proven to be a constituent, then its two conjuncts, both *two hundred* and *twenty*, must also be constituents. Some of the arguments use only multiplicative numerals as examples, as the opposite is also true, that if a multiplicative complex numeral such as *two hundred* in *two hundred books* is a constituent, then an additive complex numeral with it as a conjunct must also be a constituent, e.g., *two hundred and twenty* in *two hundred and twenty books*, under the well-accepted assumption that only constituents of the same syntactic category can be conjoined.

We now offer our final argument specifically for multiplicative complex numerals as constituents. There are numeral bases in English, e.g., *hundred*, *thousand*, *million*, and *billion*, that must combine with another numerical to be well-formed (Rothstein 2012: 528), as shown in (71).

(71) (a) One/two hundred books were sold.
      (b) *Hundred books were sold.

In fact it has been claimed as a universal that numeral bases of numeral systems are genuine one-place expressions (e.g., von Mengden 2010: 37). Rothstein (2012: 529) thus

---

[9] The two structures Danon (2012) proposes for cardinal-containing NPs in Hebrew: the cascading structure for lower cardinals two through nineteen and the specifier-head configuration for the higher cardinals thus also share the same essential feature, i.e., cardinals are constituents.

proposes that numeral bases such as *hundred* are of type <n,n> and combine with another numeral to form a complex numeral denoting the number n × 100, as in (72).

(72) ⟦ hundred ⟧= λn. {x: │ x │ = 100 × n}

If Rothstein's analysis is right, then given the phrase *two hundred books*, only the constituent account in (73), where [one hundred] is a constituent, can accommodate the syntactic and semantic requirements of *hundred* within its immediate constituent. In contrast, under the non-constituent account, such requirements of *hundred* cannot be fulfilled locally within the constituent [hundred people], as *one* is outside of this constituent.

(73) Constituent account
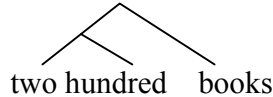
two hundred   books

(74) Non-constituent account
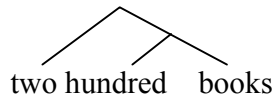
two hundred   books

What this implies is that a multiplicative complex numeral with stacking numeral bases, e.g., *two hundred thousand books*, must also have a consistently left-branching constituency structure, as in (75a). In (75b), though the entire complex numeral is a constituent, *two* and *hundred* do not form an immediate constituent, leaving the syntactic and semantic requirements of *hundred* unfulfilled. (75b) is thus ill-formed.

(75) Constituent account: two options
        (a)

two hundred thousand books

        (b)

two hundred thousand books
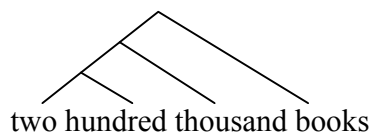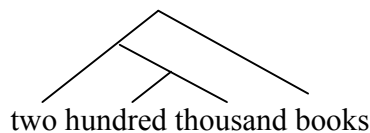
To summarize, linguistic facts from English strongly support the analysis that complex numerals in the language are syntactic constituents to the exclusion of the NP. We are unaware of any remaining arguments that IM use to motivate the non-constituency analysis that may be problematic for the constituency analysis of English numerals. However, Meinunger (2015), while recognizing complex numerals as constituents, proposes a grafting account for complex numerals cross-linguistically, thus including English. We will

now demonstrate that the grafting mechanism is entirely unnecessary for complex numerals in English.

4.5 The grafting approach

Meinunger (2015) also observes the difficulties that IM's (2006) non-constituency account suffers, centering around the inequivalence between the full version, or the source form, and the spelled-out form, i.e., the surface form, of a complex numeral, and thus advocates complex numerals as constituents cross-linguistically. Meinunger further proposes that cross-linguistically the syntactic structures of complex numerals involve 'grafting', which we will soon explain. As mentioned in the beginning of the paper, even though we support the constituency account for English, we do not believe that a particular account justified in one or some languages can be automatically extended to all languages. Meinunger first applies grafting to account for certain data in Russian, German, and Celtic that are not related to complex numerals and then extend this account for complex numerals in Russian, which is the motivation for IM's non-constituency account. Our purpose is to show that complex numerals in English are syntactic constituents plain and simple and grafting in fact complicates the grammar of English.

The grafting mechanism for the complex numeral *twenty-one* in a nominal phrase *twenty-one students* is shown schematically in (76), where & is meant to be AND, the silent counterpart of *and*. The complex numeral [twenty & one] with an upside down tree in (76c) is a separate constituent grafted upon the host sentence.

(76) (a) Complex numeral

        twenty   &   one

    (b) Host sentence

          one students came

    (c) Host sentence with graft

      twenty & **one** students came

The complex numeral in (76a) is said to be generated 'independently of the rest of the linguistic environment' (Meinunger 2015: 108). The single digit *one* in the host sentence in (76b) and the single digit *one* in the complex numeral act together like a hinge that hangs a door onto a wall. In (76c), the single digit *one* thus forms a constituent with *twenty*, {twenty

& one}, and simultaneously forms a constituent with *students*, [one students], producing the ambiguous structure {twenty & [one} students].

One motivation for this elaborate mechanism is the behavior of complex numerals in Russian. In (77), notice that, in spite of the plurality inherent in the numeral *dvacat' odin* 'twenty-one', the noun takes on a singular form, and so does the verb.

(77) Dvacat' odin   student   prišel
      twenty one   student.SG   came.SG
      'Twenty-one students came.'

This powerful mechanism of grafting thus offers a typological parameter which allows English and Russian to make different choices in terms of the marking of the grammatical feature of number. English marks the head noun as plural in accordance with the constituent {twenty & one}, while Russian abides by [one student]. Though a thorough deliberation of the cross-linguistic validity of the grafting approach to complex numerals is beyond the scope of the paper, we contend that it is unnecessary, if not unworkable, for English.

First of all, it is clear that, in terms of formal power, a grammar without the exceptional multi-dominance of structural nodes is more constrained and should be preferred. Given that complex numerals in English need to be generated as constituents, as we have argued, a nominal phrase where the complex numeral involves no single digits, e.g., *three hundred students*, does not involve grafting; the grammar thus generates (78) quite straightforwardly without ambiguity.

(78)

three hundred students

Now consider *twenty-two students*, where the single digit is *two*, not *one*. Logically, the grammar still produces two structures, one without, and the other with, grafting, as in (79a) and (79b), respectively. Given the necessity of the straightforward (78), the same structure in (79a) comes at no cost whatsoever; (79b) is thus superfluous and must be explicitly blocked to avoid such structural ambiguity.

(79) (a)

twenty & two students

      (b)
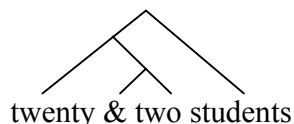
twenty & two students
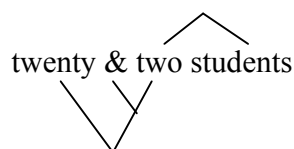
This brings us back to the graft account of *twenty-one students*. In English, unlike Russian, the behavior of *twenty-one* is exactly like other complex numerals, e.g., *twenty-two* and *three hundred*. The grammar of complex numerals in English thus enjoys better descriptive insights without grafting.

5   CONCLUSION

The central issue dealt with in this paper is whether complex numerals in English are syntactic constituents or not. The traditional view is that they are, while a recent view in the influential works by Ionin & Matushansky (2006; 2018) is that they are not. In this paper we argue for the constituent account specifically for numerals in English based on linguistic facts from English. We first demonstrate that the non-constituent account is not viable in that syntactically the underlying source forms under this account may be ill-formed and also may not be semantically equivalent with the corresponding surface forms. We further provide direct evidence from constituency tests and the behavior of post-numeral and pre-numeral approximative markers for complex numerals in English as constituents.

He (2015) demonstrates convincingly that complex numerals in Chinese are constituents, and He *et al.* (2017) further argue that complex numerals in a number of Austroasiatic, Miao-Yao, Tai-Kadai, and Tibeto-Burman languages in southern China and several Austronesian languages in Taiwan are likewise constituents. Such findings, together with those documented in this paper, pose considerable challenges to the cross-linguistic applicability of the non-constituency analysis of complex numerals. More importantly, these findings demonstrate that whether complex numerals are syntactic constituents or not in a particular language is an empirical matter and should be carefully deliberated based on linguistic facts from that language.

Authors' addresses:
Chuansheng He
Department of English and Institute of Language and Cognition
School of Foreign Languages
Hunan University
Changsha City
Hunan Province
China

One-Soon Her (Corresponding author)
Department of Foreign Languages and Literature
Tunghai University
Taichung, Taiwan
Graduate Institute of Linguistics
National Chengchi University
Taipei, Taiwan
Email: hero@thu.edu.tw

REFERENCES

Abels, Klaus & Ad Neeleman. 2012. Linear asymmetries and the LCA. *Syntax* 15(1), 25–74.

Arregi, Karlos. 2010. The syntax of comparative numerals. *Proceedings of NELS 40*. Amherst, MA: GLSA Publications.

Barwise, Jon & Robin Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4, 159–219.

Borer, Hagit. 2005. *Structuring Sense*, vol. 1: *In Name Only*. Oxford University Press. Oxford.

Carlson, Greg N. 1987. *Same* and *different*: some consequences for syntax and semantics. *Linguistics and Philosophy* 10, 531–566.

Carnie, Andrew. 2013. *Syntax: A Generative Introduction,* 3rd edn. Oxford: Blackwell Publishing.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36(3), 315–332.

Cinque, Guglielmo. 2010. *The Syntax of Adjectives: A Comparative Study*. MA: MIT Press.

Corver, Norbert & Joost Zwarts. 2006. Prepositional numerals. *Lingua* 116, 811–835.

Danon, Gabi. 2012. Two structures for numeral-noun constructions. *Lingua* 122(12), 1282–1307.

Dooley, Sheila and Ferdinand de Haan. 2008. On the nature of the approximative expression NUM-odd. In Simin Karimi, Heidi Harley, William D. Lewis, & Scott O. Farrar (eds.), *Time and Again: Theoretical and Experimental Perspectives on Formal Linguistics*, 37–53. Amsterdam: Benjamins

Franks, Steven. 1994. Parametric properties of numeral phrases in Slavic. *Natural Language & Linguistic Theory* 12, 597–674.

Giusti, Giuliana. 1997. The categorial status of determiners. In Liliane Haegeman (ed.), *The New Comparative Syntax*. Cambridge: Cambridge University Press. 94–113.

Greenberg, Joseph H. 1990. Generalizations about numeral systems. In Keith Denning & Suzanne Kemmer (eds.), *On Language: Selected Writings of Joseph H. Greenberg*, 271–309. Stanford: Stanford University Press (originally published in Joseph. H. Greenberg, Charles. A. Ferguson and Edith. A. Moravcsik, (eds.), *Universals of Human Language*, vol. 3: *Word Structure*, 249–295. Stanford: Stanford University Press, 1978).

Hackl, Martin. 2001. *Comparative Quantifiers*. Ph.D. dissertation, MIT.

Hamawand, Zeki. 2017. The notion of approximation in language. *Cognitive Semantics* 3(1), 95–120.

He, Chuansheng. 2015. Complex numerals in Mandarin Chinese are constituents. *Lingua* 164, 189–214.

He, Chuansheng, One-Soon Her, Xiaoshi Hu and Weijing Zhu. 2017. Overt coordination in additive numerals of minority languages in South China. *Syntax* 20(3), 292–316.

Her, One-Soon. 2017. Deriving classifier word order typology, or Greenberg's Universal 20A and Universal 20. *Linguistics* 55(2), 265–303.

Her, One-Soon and Hui-Chin Tsai. 2014, *Color* isn't silent, *shallow* isn't deep: Two case studies of evaluating silent elements. *Language and Linguistics* 15(6), 775–800.

Her, One-Soon and Hui-Chin Tsai. 2015. On silent elements: A case study of *grand* and its silent entourage. *Natural Language and Linguistic Theory* 33(2), 575–605.

Her, One-Soon and Hui-Chin Tsai. 2020. Left is right, right is not: On the constituency of the classifier phrase in Chinese. *Language and Linguistics* 21(1), 1–32.

Heycock, Caroline and Roberto Zamparelli. 2005. Friends and colleagues: plurality, coordination, and the structure of DP. *Natural Language Semantics* 13, 201–270.

Hofweber, Thomas. 2005. Number determiners, numbers, and arithmetic. *Philosophical Review* 114(2), 179–225.

Hurford, James R. 1975. *The Linguistic Theory of Numerals*. Cambridge: Cambridge University Press.

Hurford, James R. 1987. *Language and Number*. Oxford: Basil Blackwell.

Ionin, Tania & Ora Matushansky. 2006. The composition of complex cardinals. *Journal of Semantics* 23. 315–360.

Ionin, Tania & Ora Matushansky. 2018. *Cardinals: The Syntax and Semantics of Cardinal-containing Expressions*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1977, *X-bar Syntax: A Study of Phrase Structure*. Cambridge, MA: MIT Press.

Kadmon, Nirit. 1987. *On unique and non-unique reference and asymmetric quantification*. Unpublished PhD thesis, University of Massachusetts at Amherst.

Kayne, Richard S. 2005. A note on the syntax of numerical bases. In *Comparisons and Contrasts*. New York. Oxford University Press. 2010.

Kayne, Richard S. 2012. A note on *grand* and its silent entourage. *Studies in Chinese Linguistics* 33(2), 71–85.

Keenan, Edward. & Jonathan Stavi. 1986. A semantic characterization of natural language determiners. *Linguistics and Philosophy* 9(3), 253–326.

Krifka, Manfred. 1990. Four thousand ships passed through the lock: Object-induced measure functions on events. *Linguistics and Philosophy* 13, 487–520.

Langendoen, D. Terence. 1978. The logic of reciprocity. *Linguistic Inquiry* 9, 177–197.

Meinunger, André. 2015. Complex numeral expressions: A plea for grafts. *Syntax* 18, 103–123.

Moltmann, Friederike. 2011. Reference to numbers in natural language. *Philosophical Studies* 162(3), 499–536.

Pankau, Andreas. 2018. The structure of approximative numerals in German. *Glossa: A Journal of General Linguistics* 3(1), 21, 1–48. DOI: http://doi.org/10.5334/gjgl.325

Partee, Brabara H. 1986. Noun phrase interpretation and type-shifting principles. In Jeroen Groenendijk, Dick de Jongh & Martin Stokhof (eds.), *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. GRASS 8. Foris. Dordrecht. 115–143.

Plank, Frans. 2004. Inevitable reanalysis: From local adpositions to approximative adnumerals, in German and wherever. *Studies in Language* 28(1), 165–201.

Ritter, Elizabeth. 1991. Two functional categories in noun phrases: Evidence from Modern Hebrew. In Susan Rothstein (ed.), *Perspectives on Phrase Structure: Heads and Licensing*. In Keir Moulton (ed.), *Syntax and Semantics*, 25, 37–62. New York: Academic Press.

Rothstein, Susan. 2012. Numbers: Counting, Measuring and Classifying. *Proceedings of Sinn und Bedeutung 16*. In Ana Aguilar-Guevara, Anna Chernilovskaya & Rick Nouwen (eds.), *MIT Working Papers in Linguistics*, 527–543.

Rutkowski, Pawel. 2002. The syntax of quantifier phrases and the inherent vs. structural case distinction. *Linguistic Research* 7(1), 43–74.

Selkirk, Elisabeth. 1977. Some remarks on noun phrase structure. In Peter W. Culicover, Thomas Wasow & Adrian. Akmajian (eds.), *Formal Syntax*, 285–316. London: Academic Press.

Stavrou, Melita & Arhonto Terzi. 2008. Types of numerical nouns. *Proceedings of WCCFL 26*.

Svenonius, Pysz. 1994. The structural location of the attributive adjective. In Eric Duncan, Donka Farkas & Philip Spaelti. *Proceedings of the Twelfth West Coast Conference on Formal Linguistics*, 439–454. Stanford, CA: CSLI.

Talić, Aida. 2017. Long-form and short-form prenominal adjectives are not reduced relative clauses in Bosnian/Croatian/Serbian (BCS). *Linguistic Inquiry* 48(1), 194–212.

Vandelanotte, Lieven. 2002. Prenominal adjectives in English: Structures and ordering. *Folia Linguistica* 36(3)-(4), 219–259.

von Mengden, Ferdinand. 2010. *Cardinal Numerals: Old English from a Cross-Linguistic Perspective*. Berlin: De Gruyter Mouton.

Zweig, Eytan. 2006. Nouns and adjectives in numeral NPs. In Leah Bateman & Cherlon Ussery (eds.), *Proceedings of NELS 35*. Amherst, MA: University of Massachusetts, GLSA.