CONSTITUENCY AS A LANGUAGE UNIVERSAL: THE CASE OF LATIN

CARLO CECCHETTO – RENATO ONIGA

ABSTRACT

The main goal of this paper is to show that the flexible word order of Latin does not prevent us from fruitfully applying to Latin a set of constituency tests that have been developed in modern linguistics. Particularly, we will show that, when the concept of constituent is correctly defined, it effectively applies to Latin. We will do so by comparing English, a rigid word order language, Italian, a more flexible word order language, and Latin, which is even more flexible. This paper is organized as follows: in section 1, we offer a brief introduction on the current theoretical debate on the topic. In section 2, we argue that constituents may be formed by words that are not contiguous, at least in languages with a flexible word order. In section 3, we discuss various tests that identify VP and TP as possible constituents in Latin: these tests include pro-form substitution, ellipsis and fragment answers. Section 4 discusses tests to identify the CP layer as a constituent: these include pro-form substitution and extraposition. In the same section we deal with a possible complication arising from the use of pro-form substitution as a constituency test but we also show that it does not affect the cases we discuss. Our conclusion in Section 5 is that since constituents may be discontinuous in Latin (as in other languages), they may not be easy to identify, but they do exist.

1. INTRODUCTION

The concept of constituent is a key notion in contemporary syntactic theory. Roughly, a constituent is a group of words that functions as a natural unit within the clause. Still, the status of this concept requires

some discussion. One issue is raised by the fact that the concept of constituent has a sort of intermediate status (Carnie 2009). On the one hand, it is *not* a naïve notion, because constituents are identified by using some formal tests (typically involving displacement, ellipsis, pro-form, coordination; we return to these tests below). On the other hand, the notion of constituent is not theory internal and, as such, it is *not* formalized. For example, the typical syntax textbook starts with defining constituency tests that identify a certain group of words as a natural unit and then moves on to formally define the concept of phrase. In phrase structure grammars (say, X-bar theory in pre-minimalist versions of generative approaches or Bare Phrase Structure in more recent implementations: Chomsky 1995, 2012), a phrase is a group of words with a well-defined hierarchical structure, so a phrase is organized around a head (possibly) with a complement, a specifier and a variable number of adjuncts.

The concept of phrase is explicitly defined but is only an imperfect proxy for the concept of constituent, since there are *prima facie* cases of phrases that do not behave as constituents (and the other way around). For example, there are constituents that are smaller than phrases, as shown by the exchange in (1)-(2). The fact that the clitic *ne* can replace the group of words *orazioni di Cicerone* is standardly taken to be evidence that this group of words is a natural unit (therefore a constituent). However, *orazioni di Cicerone* is not a complete phrase in (1) but just a subpart of the quantifier phrase *cinque orazioni di Cicerone*, which includes the quantifier *cinque*.

(1)   Ho letto cinque orazioni di Cicerone.
       "(I) have read five speeches by Cicero".
       I read five speeches by Cicero.

(2)   Io ne ho lette tre.
       "(I) of.it have read three".
       I read three of them.

Another reason suggesting that it is not possible to assume a straightforward one-to-one phrase/constituent correspondence is that a phrase can form a unit (so, behaving as a constituent) before an occurrence of syntactic movement splits it apart. For example, a classical analysis of the second sentence in (3) is in term of VP-ellipsis. The fact the verb and the direct object can be elided shows that these two words are a constituent, namely a verb phrase (VP). However, in contemporary syntactic theory it is assumed that the subject is generated inside the VP before moving to its dedicated position (Spec,T). So, one can ask why the subject is not deleted in the second clause in (3), since the VP is. The answer is that, by time the VP is deleted, the subject has already escaped the ellipsis site (leaving there a trace that would not be pronounced even without ellipsis).

(3)   John bought a house and Mary$_i$ did t_i ~~buy a house~~ too.

In English, which is a rigid word order language, it is relatively easy to track the occurrences of syntactic movement that can disrupt the phrase/constituent parallelism. And in fact, for English, it is tacitly (and sometimes explicitly) assumed that constituents are formed by *contiguous* words. However, the situation is much more difficult (and therefore more interesting) in languages that have a more flexible word order. A paradigmatic case of such a language is Latin, where constituent identification is admittedly harder. When confronted with this situation, there are two possible approaches. The first one, universalist in spirit, is to look for constituents even when they may be difficult to identify. If one takes this view, one should focus on those constituency tests that are less dependent on linear order and critically examine them. The second approach is arguing that languages with 'free' word order are radically different and that the notion of constituent does not apply to them. This anti-universalist view is explicitly endorsed by Evans – Levinson (2009), who take Latin to be a chief counterexample to the idea that the concept of constituent is a language universal. In fact, Latin becomes an important argument

supporting their claim that language universals do not exist and are even a perilous 'myth' (Evans – Levinson 2009: 440-442).

We believe that it is methodologically wrong to postulate a fundamental difference between languages without previously attempting unification. In fact this attitude might even be considered anti-scientific, since science should look for unification under superficial differences. For this reason in this paper, we try to do what Evans – Levinson (2009) should have done, namely apply to Latin constituency tests initially developed for languages with a more rigid order. To be sure, their application to Latin is not straightforward and a series of important theoretical questions arise. The facts that new research questions emerge however, far from being a reason to abandon a framework for investigation, shows its strength. More particularly, we will show that, when the concept of constituent is correctly defined, it fruitfully applies to Latin. We will do so by comparing English, a rigid word order language, Italian, a more flexible word order language, and Latin, which is even more flexible.

2. CONSTITUENTS DO NOT NEED TO BE FORMED BY CONTIGUOUS WORDS

What is a constituent? Given what we said above, namely that the concept is not formalized, answering this question means defining a set of necessary or sufficient conditions to be met for a group of words to be a constituent. Classical constituency tests qualify as *sufficient* (not necessary) conditions. Let us take a concrete example, by using the movement test: «if a group of words can be moved from its base position to some other position in the sentence, that group of word qualifies as a constituent in that sentence». Given the movement test, *that book* is a constituent in (4):

(4)   [That book]$_i$ I read t_i (not this one).

Is passing the movement test a necessary condition for constituent-hood? No, it is not. Consider another classical constituency test, namely coordination: «if two groups of words can be conjoined by an element like *and*, *or* and *but*, each conjoint is a constituent». Given the coordination test, *to the president* and *to his lawyer* are two constituents in (5):

(5)   John sent his memoir to the president and to his lawyer.

However, there is no grammatical way to build a grammatical sentence by moving the constituent *to his lawyer* alone. For example, (6) is totally ungrammatical:

(6)   *[To his lawyer]$_i$ John sent his memoir to the president and t_i.

So, *to his lawyer* passes the coordination test but does not pass the movement test (the ungrammaticality of (6) can be attributed to Ross' (1967) Coordinate Structure Constraint, which says that you cannot extract a conjoint by leaving the other one *in situ*). Still, *to his lawyer* is a constituent in (6), showing that constituency tests are sufficient (not necessary) conditions. In fact, the complicated business of having a number of different constituency tests is motivated by this. If constituency tests were necessary and sufficient conditions, one test would be enough.

Let us ask if adjacency is a sufficient or necessary condition for words belonging to the same constituent. That adjacency is not a sufficient condition is obvious, as shown in the first pages of any syntax textbook. For example, *sent his* in (5) is not a constituent (it cannot be moved, it cannot be elided, it cannot be replaced by a pro-form etc.). Is adjacency a necessary condition? If one considers only rigid word order languages like English, one might be tempted to answer "yes" because, as we mentioned, constituents are normally formed by *contiguous* words. However, it is not necessary to go as far in the scale of word-order flexibility as Latin to understand that there

are constituents which are formed by non-contiguous words. Take sentence (7) in Italian. A standard constituency test is pro-form substitution (if a group of words can be replaced by a pro-form, then that group of words is a constituent). In Italian, a special case of pro-form is the verbal anaphor *farlo* "to do that". This pro-form identifies the sequence of words *mangia... la minestra* "eats… the soup" as a constituent in (7), although it is discontinuous (it is interrupted by the adverb *con piacere*). Note that the pro-form does not replace the bigger constituent *mangia con piacere la minestra*, as the second sentence in (7) would be contradictory (Maria cannot eat the soup gladly *and* against her will)[1].

(7)   Gianni mangia con piacere la minestra, invece Maria lo fa controvoglia.
      "Gianni eats gladly the soup, but Maria it does against-her-will".
      John eats the soup gladly, but Mary does that against her will.

Note that our main point, namely that constituents may be discontinuous, holds even if one assumes that *fare* is a place holder for tense and agreement features of the verb (much like *do* in English *do*-support construction) while the pronoun *lo* is a pro-form for the uninflected verb plus the remaining categories in the VP. Assuming this analysis for *farlo*, our observation can be reformulated by saying that it is the pro-form *lo* that identifies the sequence *mangia… la minestra* as a discontinuous constituent in (7).

We took Italian as an example, but (superficially) discontinuous constituents are in fact the norm in most languages, rigid word order languages like English being the exception rather than the norm. So, we can conclude that adjacency is neither a sufficient nor a necessary condition for constituency. We stress this point because Evans –

---

[1] If the semantics allows it, *farlo* may substitute the bigger constituent [mangia con piacere la minestra] as in (i):

(i)   Gianni mangia con piacere la minestra e anche Maria lo fa.
      "Gianni eats gladly the soup and also Maria it does".
      John eats the soup gladly, and Mary does too.

Levinson (2009: 440) claim that constituent structure cannot be a language universal because «many languages show few traces of constituent structure, because they scramble the words». Scrambling of words, which creates discontinuous constituents, is an interesting property and deserves a close examination. However, since constituent identification is done on the basis of tests that may apply to words that are not adjacent, scrambling of words is no special challenge for the hypothesis that constituent structure is a language universal.

Returning to Latin, it has been argued (Devine – Stevens 2006; Giusti – Oniga 2007; Iovino 2011; Ledgeway 2012) that even extreme cases of "scrambling of words" in Latin show a certain regularity and their distribution can only be understood by assuming the presence of a constituent structure.

3. CONSTITUENCY TESTS IN LATIN

Having established that word order flexibility is not a principled obstacle to constituent identification, let us look for suitable constituency tests in Latin.

3.1. VP ELLIPSIS AND TP ELLIPSIS

Ledgeway (2012: 194) argues that it is possible to identify VP ellipsis in Latin, so, if he is right, VP is a genuine constituent in Latin. He mentions cases such as (8) in which an infinitival complement of the modal predicate *volo* "want", although realized in the *a.* sentence, is absent, though understood, in the *b.* sentence. Similarly, the infinitival complement of *possum* "can" is realized in the *c.* sentence but is deleted in the *d.* sentence:

(8)   Plaut. *Poen*. 308-9

a.   Agorastocles: Eho tu, vin tu facinus facere lepidum et festivom?
      "hey you.VOC you.want.Q you.NOM deed.ACC do.INF witty.ACC
      and merry.ACC".
      How now, you; do you want to play a merry and a frolicsome prank?

b.   Milphio: Volo.
      "I.want".
      I do.

c.   Agorastocles: Potesne mi auscultare?
      "you.can.Q me.DAT listen.to.INF".
      Can you, then, give attention to me?

d.   Milphio: Possum.
      "I.can".
      I can.

We concur with Ledgeway that the examples in (8b) and (8d) are clear instances of ellipsis. We also assume that ellipsis is phonological deletion of syntactically fully fledged structure and this deletion takes place on the PF-branch of the syntactic computation[2].

However, it is less clear whether the elided constituent is a VP or a bigger constituent. In canonical generative approaches, the representation of the clause includes a CP layer, the syntactic level which encodes information about sentence force (declarative, interrogative, imperative), the subordinate/matrix character of the sentence, the finiteness/non finiteness distinction, and other relevant syntactic information. CP selects for the TP layer, which in turn selects for the VP layer. We go back to cases of CP ellipsis in Latin in Section 4. Our question now is whether the elided constituent in (8) is a VP or a TP[3]. In a previous work of ours (Cecchetto – Oniga 2002),

---

[2] See Merchant (2001) and Aelbrecht (2010), among many others, for an account along this line for ellipsis cases.

[3] For the sake of simplicity in this paper, as in Oniga (2014), we will stick to the traditional partition of clause structure in the VP, TP and CP areas. Assuming the

we have argued that *possum* and *volo* are different in this respect, because the former, but not the latter, takes a VP structure (not a TP) as a clausal complement. The main empirical reason for assuming this difference is that, although Latin infinitives are fully fledged for tense (there is a morphological form for present, past and future tense both in the active and in the passive voice), *possum* never selects for an inflected infinitive. This is shown by the contrast between the grammaticality of (9), where there is the default infinitival form (which morphologically corresponds to a present infinitive), and the ungrammaticality of (10) and (11), where the infinitive is inflected for past tense (10) and for future tense (11). Note that the ungrammaticality of (10) and (11) is not due to semantics in any obvious sense, since the propositional content "It is possible for me to have read/to read in the future" is coherent.

(9) Sen. *epist.* 33, 9

    Legere possum.
    "read.INF.PRES I.can".
    I can read.

(10) *Legisse possum.
    "read.INF.PAST I.can".

(11) *Lecturus esse possum.
    "read.INF.FUT I.can".

---

rich set of functional projections postulated in the cartographic tradition (Rizzi 1997 and Cinque 1999 among many others), the constituency tests that we are going to discuss should be further scrutinized to see which portion of these areas undergo the relevant phenomena. We leave this to future work.

Interestingly, the verb *volo* is different, since it can select for a past infinitive, as shown by the grammaticality of (12)[4].

(12) Cic. *Att.* 14, 18, 4

> Legisse vellem.
> "read.INF.PAST I.would-like".
> *I would like to have read.*

We take the fact that an infinitive inflected for tense is not admitted as complement of *possum* to be evidence that the T layer cannot be projected in the complement clause of *possum*. This means that the complement of *possum* in (8c) (*mi auscultare*) is a VP, not a TP. So, we can safely conclude that (8d) contains a genuine case of *VP* deletion. As for the complement of *volo* in (8a) (*facinus facere lepidum et festivom*) there is no absolute certainty that it is a VP, as *volo* in principle can take a TP as clausal complement. So, the ellipsis case in (8b) might be either a case of VP deletion or a case of TP deletion. We leave this question open.

Be that as it may, following Ledgeway (2012), we have identified at least a clear case of VP ellipsis in Latin. As VP ellipsis is a paradigmatic constituency test, we have started gathering evidence for a constituent structure in Latin.

However, a legitimate question is whether the only attested cases of ellipsis involve clausal complement of modal verbs. We think that the phenomenon is more general. In Section 3.2 we use another constituency test to show cases of ellipsis not involving modal verbs.

Here we mention that it is tempting to analyse (some of) the so-called 'verb sparing' cases (Hofmann – Szantyr 1965: 423. 823) as VP

---

[4] The difference between *volo* and *possum* is reflected in the fact that only the former admits the *Accusativus cum Infinitivo* construction (see Cecchetto – Oniga 2002 for discussion).

ellipsis, since not only the verb but also its arguments may be spared ('elided' in modern terms)[5].

   An interesting case of ellipsis is found in (13):

(13) Cic. *Phil.* 2, 61

> Si te municipiorum non pudebat, ne veterani quidem exercitus? (*scil. te pudebat*)
> "if you.ACC of.the.citizens not feel.shame.IMPF, not of.the.old even of.the.army".
> *If you did not feel shame for the citizens, didn't you* (implied. *feel shame*) *at least for the army of veterans?*

   In (13) the elided constituent (*te pudebat*) has a linguistic antecedent, much like canonical ellipsis cases. Incidentally, (13) illustrates a case of discontinuous constituent, as the words *te* and *pudebat*, which form the elided constituent, are not contiguous in the protasis. One issue that remains open is how it is possible that *veterani exercitus* is not elided, since it is an argument of the verb *(te) pudebat*. A natural analysis might be in terms of TP ellipsis, as suggested by

---

[5] Goldberg (2005) distinguishes two types of languages as far as VP ellipsis is concerned. The first type (Aux-Stranding VP ellipsis languages) is exemplified by English. In English VP ellipsis, the verb is elided together with other VP internal material ('John broke a vase, and Mary did broke a vase too'). This happens because no V-to-I movement occurs in English (Pollock 1989), so, when the VP is deleted, the verb is also deleted. The second type (V-Stranding VP ellipsis languages) is exemplified by Hebrew, Irish, and Swahili. In these languages a finite verb survives VP ellipsis, while all other categories in the VP are elided. So a sentence which is a word by word translation of "John broke a vase, and Mary broke a vase too", which is ungrammatical in English, is fully acceptable in these languages. This happens because the verb moves out of the VP, so, when the VP is deleted, the verb survives ellipsis. Given this very general typology, and assuming that the verb, given its morphological richness moves out of the VP in Latin, one expects Latin to be a V-Stranding VP ellipsis language, therefore verb sparing cases are expected. However, Latin is still different, since, given its scrambling properties, not only the verb but a subset of the arguments of the verb can be spared.

Danckaert (2012: 92), who applies to Latin an analysis originally proposed by Merchant (2003) for so-called stripping in English[6]. (13) would be analysed as follows: first *veterani exercitus* moves to a left peripheral position and leaves behind a TP with a gap (we assume that *quidem* sits in the CP layer). Subsequently, this TP is phonologically elided in the PF component. The constituent *veterani exercitus* survives deletion, since it has moved to a position higher than the ellipsis site. Adopting this perspective, (13) would be a case of TP ellipsis.

Another class of cases amenable to an ellipsis analysis is the *apò koinoû* construction in Plautus (Leo 1896: 44), illustrated in (14).

(14) Plaut. *Rud.* 1043

Si adhibebit fidem, etsi ignotust, notus, si non (*scil.* adhibebit fidem), notus ignotissimumst.
"if he.will.apply faithfullness.ACC, even.if unknown.is, known.NOM, if not, known.NOM very.unknownNOM".
*If he will be faithful, even if he is unknown, he will be known, if he will not (*implied. *be faithful), even if he is known, he will be very unknown.*

Even in this case, the natural question that arises is whether the elided constituent (*adhibebit fidem*) is a VP or a TP. Answering this question would require an analysis of the position of the negative head *non* in Latin. If *non* heads a NegP phrase in the TP area, (14) should be analysed as a case of VP deletion. However, Zanuttini (1997) has argued that many Romance varieties have a position for negation higher than the TP area. If Latin *non* admits a similar analysis, ellipsis

---

[6] Stripping (Hankamer – Sag 1976: 409 for the initial definition) is a type of ellipsis under which everything in a clause is deleted under identity with corresponding parts of the preceding clause, except for one constituent and (usually) an adverb or a negative element. Sentence (i) minimally contrasts with the VP ellipsis case in (ii), because in (i) also the T/INFL node is deleted.
(i)   John broke a vase, and Mary too.
(ii)   John broke a vase, and Mary did too.

(14) is compatible with a TP deletion analysis. Since our main goal in this paper is offering arguments for the existence of constituent structure in Latin, we remain agnostic on whether TP or VP deletion occurs in (14).

In the next section we move to another type of constituency test.

3.2. FRAGMENT ANSWERS AND TP ELLIPSIS

Another classical constituency test is the fragment answer test. A popular conception is the following: the only fragments that are allowed as answers are those that would form a constituent in the complete answer.

For example, the fragment answer in (15a) is acceptable because *a friend of mine* is a constituent in the complete (if redundant) answer *I have met a friend of mine*. The fragment answer in (15b) is not acceptable because "have met a" is not a constituent in the complete answer.

(15) Who have you met?
    a. A friend of mine.
    b. *have met a.

So, fragmentary utterances are interpreted as fully sentential structures but for the fact that ellipsis occurs in the PF component. For example, (15a) should be analysed as in (16). First, the object is moved to a position in the CP area. Following this movement, the TP where the trace of the object is found gets phonologically elided (this analysis is similar to the one considered above for (13)).

(16) [$_{CP}$ [A friend of mine]$_i$ [$_{TP}$ ~~I have met t$_i$~~]]

To sum up: fragment answers can be analysed as evidence of TP ellipsis, therefore as evidence that TP is a constituent in a given

language. This reasoning rather straightforwardly applies to Latin. (17) to (20) are all cases where the understood (and elided) constituent can be analysed as a TP containing a trace of the fragment. The fragment survives phonological deletion because it has moved out of the elided constituent: we show this explicitly for (17) in (17'):

(17)   Plaut. *Men.* 285-6

   Menaechmus: Quem tu parasitum quaeris, adulescens, meum?
   "what.ACC you hanger-on.ACC you.look.for, guy.VOC, my.ACC".
   Which hanger-on of mine are you looking for, guy?

   Cylindrus: Peniculum (scil. ego quaero).
   "Peniculus.ACC".
   Peniculus (implied: *I am looking for).*

(17')   [$_{CP}$ Peniculum$_i$ [$_{TP}$ ~~ego quaero t$_i$~~]]

(18)   Plaut. *Amph.* 717

   Sosia: Tun' heri hunc salutavisti?
   "you.NOM.Q yesterday him greet.PAST".
   Did you greet him yesterday?

   Alcumena: Et te quoque etiam, Sosia!" (*scil.* ego salutavi)
   "and you.ACC also too Sosia.VOC".
   And you too, Sosia! (implied: *I greeted).*

(19)   Plaut. *Amph.* 391

   Sosia: Tuae fidei credo?
   "your.DAT faithfulness.DAT I.believe".
   Do I believe your words?

   Mercurius: Meae (*scil. fidei crede*).
   "my.DAT".
   Mine (implied: *I believe).*

(20) Plaut. *Amph*. 450

> *Mercurius*: Quo agis te?
> "where you.lead you.ACC".
> *Where are you going?*
>
> *Sosia*: Domum (*scil.* ago me).
> "home.ACC".
> *Home* (implied: *I am going).*

One might ask what type of movement is responsible for removing the fragment (say *Peniculum* in (17)) from the ellipsis site. Given that Latin has a very flexible word order, it has various scrambling operations that can be deemed responsible for this movement. In particular, a possibility is focalization, since the fragment answer corresponds to new information. When the movement that lets the fragment survive is due to focalization, the Latin cases resemble sluicing in English, for example sentences like *Someone arrives but I do not know who*[7].

However, since many more scrambling operations are available in Latin than in English, more categories may survive TP ellipsis in Latin[8]. In fact, a typical way to answer *yes/no* questions in Latin is to repeat the main verb of the question (Brown *et al.* 2009). This strategy is often employed cross-linguistically and has been interpreted as realizing polarity focus in C (Holmberg 2007 for a distinction between languages that standardly reply to yes/no questions by repeating the finite verb of the question and languages that do so by using a special affirmation particle "yes").

However, there are a few cases in which the movement that allows a category to survive ellipsis might not be due to focus. We mention one in (21):

---

[7] Merchant (2001), Merchant – Simpson (2012) and Marušič – Žaucer (2013).
[8] See footnote 5.

(21) Plaut. *Amph.* 607

> *Amphitruo*: Quis te verberavit?
> "who you.ACC he.has.beaten".
> *Who beat you?*

> *Sosia*: Egomet memet (*scil. verberavi*).
> "I.self.NOM my.self.ACC.
> *I* (implied: *have beaten) myself.*

In (21) the fragment is composed of two categories (*ego*, the nominative first person pronoun*,* and *me*, the accusative first person pronoun, both reinforced by the suffix *-met* "self") which do not form a constituent under standard assumptions (it is uncontroversial that subject and object do not form a constituent that excludes the verb). So, *egomet* and *memet* must have moved one by one out of the ellipsis site, as illustrated in (21').

(21') [$_{CP}$ Egomet$_i$ memet$_j$ [$_{TP}$ t$_i$ ~~verberavi~~ t$_j$]]

The structure in (21') is not easily amenable to a simple case of focalization, as it would be a case of multiple foci and independent movement of multiple foci has been claimed to be impossible (Rizzi 1997). It is possible that the movement of the two categories is of different type, for example one is a focus movement while the other is a topic movement[9]. That topic and focus can co-occur in principle in a fragment answer is shown in the Italian example (22):

---

[9] (21) resembles in some respect cases of gapping. Gapping, exemplified in (i), is a type of ellipsis in which the finite verb in the second conjunct of a coordinated structure is omitted under identity with the verb in the first conjunct.
(i)   Some like football, and others baseball.

(22) *Question*: Chi vi ha scritto?
       Who to-you has written.
       Who wrote to you?

       Answer 1:   A me, Gianni.
                       To me Gianni.

       Answer 2:   E a me, Maria.
                       And to me Maria.

In this section we have shown that fragment answers offer an indirect, but interesting, argument for the constituenthood of the TP in Latin. Since we have previously considered evidence for the constituenthood of the VP, the picture that emerges from applying standard constituency tests is that Latin, despite its flexible word order, displays the hierarchical structure that is considered a landmark of any natural language by generative approaches. Flexible word order should not prevent us from applying constituency tests and when we do that, the result is the familiar clause skeleton with a VP layer being embedded into a TP layer.

Let us conclude with a *caveat* concerning the methodological problem that arises with languages which do not have native speakers anymore. While in 'living' languages one can show, by means of grammaticality judgments, that ellipsis is a test for constituency in that we cannot delete any string, for Latin there is no direct information from native speakers as to what is *not* possible. In principle, a skeptical reader might argue that we have just shown that deletion of linguistic material is possible in Latin, but maybe any string could be deleted. However, it is likely that the Latin corpus is large enough to be a proxy (although an imperfect proxy) of the actual varieties that were spoken. In addition, a rigid application of the skeptical view would impede the study of 'dead languages' altogether, since any conclusion that is reached on the basis of the available evidence in principle might be reversed by accidental gaps in the corpus. So, while the skeptical objection cannot be easily dismissed, we think that it

should not prevent us from forming hypotheses on the basis of the available evidence.

3.3. VERBAL ANAPHORA

As mentioned, many languages have special cases of pro-forms which roughly correspond to the verbal anaphors "to do that" and "to do so" in English. The Latin verbal anaphor *idem facere* is a relevant case.

Under the standard assumption that, if a group of words can be replaced by a pro-form, then that group of words is a constituent, we can identify as constituents the groups of words replaced by the inflected forms of *idem facere*. These are in italics in the examples (23)-(25).

(23) Curt. 10, 7, 19

Primus Perdicca *arma deposuit*, ceterique idem fecere.
"first.NOM   Perdicca.NOM   weapons.ACC   put.down.PAST, others.NOM.and the.same.ACC did".
Perdicca was the first one to put down the weapons, and the others did the same.

(24) Caes. *Gall.* 1, 15, 1

Postero die *castra ex eo loco movent*. Idem facit Caesar.
"Following.ABL day.ABL camp.ACC from that.ABL place.ABL they.move. the.same.ACC does Caesar.NOM".
The day after they moved the camp from that place. Caesar did the same.

(25) Amm. 16, 12, 35

> Chonodomarius *iumento statim desiluit* et secuti eum residui idem fecere.
> "Chonodomarius.NOM horse.ABL immediately down.jumped and following.NOM him others.NOM the.same.ACC did".
> *Chonodomarius immediately got off his horse, and the others did the same following him.*

The usual question arises at this point: is the constituent replaced by *idem facere* a VP or a TP? The simplest analysis, which we adopt, is assuming that it is a VP. There is no indication that the subject occupies a special position in the left periphery in (23) to (25). In fact, there is no indication that the left periphery layer is filled by any lexical material in these sentences. It is worth comparing sentences in (23) to (25) with (13) and (14) above, where the lexical items *quidem* and *non* might be analysed as sitting in the CP layer. The elliptical clause in (23) to (25) seems to be a garden variety TP structure. If so, given that the subject is not replaced by a verbal anaphor, we can conclude that the pro-form substitution test allows us to identify *VP* constituents in Latin.

4. PRO-FORMS FOR THE CP LAYER

As we already mentioned, in canonical generative approaches, the representation of the clause includes a CP layer, which encodes information about illocutionary force (declarative, interrogative, imperative), the subordinate/matrix character of the sentence, the finiteness/non finiteness distinction and other relevant syntactic information. CP includes the TP layer, which in turn includes the VP layer. There is evidence for the existence of a CP layer in Latin. In fact, it is possible to show that this layer includes information that in other languages is not easy to associate with the CP, for example information about polarity (affirmative/negative), as shown by the

alternation between the complementizer *ut* "that" and the complementizer *ne* "that not".

But is it possible to identify the constituent corresponding to the CP layer by using constituency tests in Latin? One revealing case is (26), in which *idem* is a pro-form for the embedded clause *pauperorum filias ut indotatas ducant uxores domum* "that they marry the daughters of poor men with no dowry".

(26) Plaut. *Aul.* 478-81

Nam meo quidem animo, si idem faciant ceteri / opulentiores, pauperiorum filias / ut indotatas ducant uxores domum / et multo fiat civitas concordior.
"indeed my.ABL at.least opinion.ABL, if the.same.ACC do.CONJ other.NOM.PL   more.rich.NOM.PL   more.poor.GEN.PL daughter.ACC.PL   that   without.dowry.ACC.PL   lead.CONJ wives.ACC home.ACC and much would.become town.NOM more.harmonious.NOM".
And at least in my opinion, if the other rich men did the same, namely if they married the daughters of poor men with no dowry, the town would be more harmonious.

(27) and (28) are slightly different, as we see a pro-form (*idem facere*) which replaces a complex CP composed of a matrix and a subordinate clause. In (27) *idem facere* is a pro-form for *cohortis paulatim incedere iubet* ("orders the units to advance slowly"). In (28) where *idem facere* stands for *iurarent se nisi uictorem in castra non reuersurum* ("swore not to go back to the camp unless victorious"), while in (29) it stands for *itinere facto consedit* ("after marching, he stopped").

(27) Sall. *Cat.* 60, 1

Petreius tuba signum dat, cohortis paulatim incedere iubet; idem facit hostium exercitus.
"Petreius.NOM trumpet.ABL signal.ACC gives units.ACC slowly advance.INF orders the.same.ACC does enemies.GEN army.NOM".
Petreius gave a signal with the trumpet and ordered the units to advance slowly; the army of the enemies did the same.

(28) Caes. *civ.* 3, 87, 5

Haec cum dixisset iuravit se nisi victorem in castra non reversurum reliquosque ut idem facerent hortatus est.
"that.ACC.PL after he.say.CONJ.PLPF he.swored himself if.not victorious.ACC in camp.ACC non go.back.INF.FUT others.ACC. and that the.same.ACC they.do.CONJ.IMPF he.exhorted".
After having said that, he swore that he would not go back to the camp unless victorious, and exhorted the others to do the same.

(29) Sall. *Iug.* 91, 3

Dein postquam tempus visum, castris egreditur, noctemque totam itinere facto consedit; idem proxuma facit.
"then after time.NOM it.seemed camp.ABL go.out night.ACC.and all.ACC march.ABL done.ABL he.stopped the.same.ACC next.ABL he.does".
Then, when he thought that it was time, he went out from the camp and, after marching all the night, stopped; he did the same the following night.

Notwithstanding their importance, verbal anaphora tests, as the one we are currently considering for Latin, are affected by a potential problem that is worth discussing. We illustrate this problem by using the English pro-form "to do that":

(30) John bought a book before Mary did.

(31) John bought a book before Mary did that.

(30), which is a case of VP ellipsis, is not ambiguous: it can only mean that John bought a book before Mary bought a book. However (31) is ambiguous: it can either mean that John bought a book before Mary bought a book or that John bought a book before Mary did some other action (perspicuous in the context of utterance). This introduces a complication: while VP ellipsis in sentences like (30) requires a linguistic antecedent (the matrix VP "buy a book"), a verbal anaphor does not always need a linguistic antecedent. If a meaning for the verbal anaphor is available, it can be correctly interpreted even if it corresponds to no *specific* linguistic antecedent. Notice that in principle even the reading of (31) in which the sentence means that John bought a book before Mary bought a book might be obtained even if "to do that" did not take a linguistic antecedent. For example, the main clause might shape the context of utterance, and by doing so it might make available a suitable meaning for the "to do that" anaphor. So, the link between the verbal anaphor and its linguistic antecedent would be indirect even in (30). For this reason, it can be argued that ellipsis tests are more reliable than verbal anaphora tests.

Given this general proviso, can we really be sure that Latin anaphors like *idem* or *idem facere* require a linguistic antecedent? Notice that, if indirect licensing of *idem* and *idem facere* by its antecedent suffices (much like in the case 31, just discussed), we would not have a strong argument for the constituenthood of the CP in (26) to (29). Similar considerations would apply to other uses of verbal anaphora as a constituency test. However, we think that we can fix this potential problem. The first observation concerns specifically *idem facere*. The expression *idem* "the same", by virtue of its lexical meaning, limits the possibility of an extra-linguistic antecedent, as suggested by the fact that the only meaning of (31') is "John bought a book before Mary bought a book".

(31') John bought a book before Mary did the same.

Similarly, the only meaning available in (26) to (29) is the one that we pointed out above, where *idem* is interpreted as having the same meaning as the antecedent CP.

A more general consideration in favour of the reliability of verbal anaphora as a constituency test builds on the observation that even pro-forms that in principle can take a non-linguistic antecedent are not always allowed to do that. This can be clarified with an example. Many Italian speakers accept a transitive use of *stramangiare* "over-eat". For those who do, the second sentence in (32) *Piero ha fatto la stessa cosa in piccole dosi*, either is not an acceptable continuation at all (because it is weird to overeat by ingesting small quantities) or, as long as it is acceptable, can only mean that Piero has overeaten too, although he has done that by eating many small quantities of cake in a row. This contrasts with (33), in which exactly the same sentence *Piero ha fatto la stessa cosa in piccole dosi* does not presuppose that Piero has overeaten. In fact, the most natural interpretation of (33) is that Piero has *not* overeaten.

(32) Gianni ha stramangiato la torta. Piero ha fatto la stessa cosa in piccole dosi.
      Gianni has over-eaten the cake. Piero has done the same thing in small quantities.

(33) Gianni ha mangiato la torta in grandi dosi. Piero ha fatto la stessa cosa in piccole dosi.
      Gianni has eaten the cake in big quantities. Piero has done the same thing in small quantities.

If the verbal anaphora *fare la stessa cosa* "to do the same thing" requires a linguistic antecedent, we can easily account for the contrast between (32) and (33). The only available linguistic antecedent in (32) is the VP *stramangiare la torta* "over-eat the cake", and this explains why Piero must have overeaten (even if it is strange to over-eat by ingesting little quantities). In (33) the linguistic antecedent can be the VP *mangiare la torta* "eat the cake", since the idea that Gianni ate a

lot is expressed by the independent PP *in grandi dosi* "in big quantities".

On the other hand, suppose that the meaning of the verbal anaphor *fare la stessa cosa* could be retrieved from the extra-linguistic context. Since the first sentences in (32) and (33) are roughly synonymous, no contrast in interpretation should arise between (32) and (33): whether modification of the verb meaning is expressed by a PP (33) or by a prefix on the verb (32) should not be important, as long as the activity of overeating is made contextually salient by uttering the first sentence in each pair.

Therefore, the generalization that emerges is that a pro-form cannot take a non-linguistic antecedent if a linguistic antecedent is available that conveys the same meaning as the non-linguistic one.

As expected, this generalization is not limited to Italian. For example, Jason Merchant (p.c.) pointed out to us examples in English that suggests that this generalization might be valid crosslinguistically. While (35) is fully acceptable, (34) is odd for the same reason that is (32) is: in (34) the linguistic antecedent cannot be ignored.

(34) *John retook the exam and Mary did that for the first time.

(35) John took the exam for the first time and Mary did that for the second time.

Crucially, in all the sentences in (26) to (29) a linguistic antecedent *is* available and this is the CP that we indicated above. So, assuming the cross-linguistic validity of the generalization that a linguistic antecedent cannot be ignored when present, we can conclude that in (26) to (29) the pro-form takes a linguistic antecedent, since it can.

A final consideration can support the conclusion that *idem* requires an antecedent in cases like (26) above. Superficially (26) resembles cases of cataphora, since *idem* precedes the CP *pauperorum filias ut indotatas ducant uxores domum*. So, one cannot say that *idem* receives its meaning from the context set up by the *preceding* discourse. More

precisely, (26) seems to contain a syntactic dependency between the *ut*-clause and *idem*, as in English cases of extraposition like (36), where there is syntactic dependency between a CP extraposed at the end of the sentence and the pronoun "it" in the canonical position of the CP (Rosenbaum 1967 and much following work):

(36) a. I could not believe **it** for a second **that he failed the exam**.
       b. **It** surprises me **that doctor came at all**.

Given the presence of a syntactic dependency between the pro-form and a CP, it would be strange if the interpretation of the pro-form (*it* in English or *idem* in Latin) were not syntactically mediated. Notice that, as expected, a dependency involving an extraposed CP can occur with pro-forms different from *idem* in Latin. In (37), for example, the pronoun *illud* ("that") is linked to extraposed CP *ut philosophiam in praecordia ima demittas* ("that you let philosophy go deep in your heart").

(37) Sen. *ep.* 20, 1

      Illud te rogo atque hortor, ut philosophiam in praecordia ima demittas.
      "That you.ACC I.beg and I.exhort that philosophy.ACC in heart.ACC
      deep.ACC you.let.go.CONJ".
      I beg and exhort you, that you let philosophy go deep in your heart.

Let us take stock: there are clear cases of long distance dependencies between a pronoun and an extraposed CP and for these dependencies it is only natural to think that the meaning associated to the pronoun is linguistically mediated, namely the pronoun receives its meaning directly from the CP it is linked to.

We conclude this section with a caveat. Of course from what we said it does not follow that there may not be cases where *idem facere* (or another pro-form) receives its meaning from the general context of utterance rather than from a specific constituent in the sentence. In

fact, it is not difficult to find cases like that. An interesting example is (38):

(38) Plaut. *Stich*. 641-44

> More hoc fit atque stulte, mea sententia: / Si quem hominem exspectant, eum solent provisere, / Qui, hercle, illa causa ocius nihilo venit. / Idem ego nunc facio, qui proviso Sagarinum.
> *It is always like that, and stupidly, in my opinion: if you are waiting for someone, you head for him, but – by Hercules – he will not arrive any sooner because of this. Now I do the same, while I am heading for Sagarinus.*

By uttering this sentence, the speaker is saying that, although he knows that heading for someone does not accelerate his arrival, he will (stupidly) head for Sagarinus nonetheless. So, in this case *idem facere* seems to refer to the complex propositional content set up by the preceding discourse. This propositional content can be paraphrased as follows: "waiting for someone and stupidly going out to meet him although this does not make him arrive sooner". Since this complex content does not match a single syntactic constituent, *idem facere* does not correspond to a constituent in (38).

Interestingly, this double life of *idem facere* was already observed in traditional grammars of Latin for the phenomenon of ellipsis, which is said to be the omission of *Begriffe oder Satzteile*, that is "concepts or parts of the discourse" (Kühner – Stegmann 1976: 549). In a similar way, we can oppose "Satzteile pro-forms" (linguistically controlled pro-forms in modern terms) and "Begriffe pro-forms" (pro-forms whose meaning is a propositional content which does not necessarily correspond to a single syntactic unit). So, we are not inventing anything new here. However, we may have found a more precise methodology to set the two different types of pro-form apart.

In this section, we have shown that there are uses of pro-forms which qualify as reliable constituency tests. By applying these tests we have identified the CP as one structural level of the clause in Latin.

5. CONCLUSION

Our main goal in this paper was showing that the flexible word order of Latin does not (and should not) prevent us from fruitfully applying to Latin the set of constituency tests that have been developed in modern linguistics. Contrary to the Evans – Levinson's (2009) claim that Latin would not have constituents, we have discussed evidence that the VP, the TP and the CP can all be identified by constituency tests, most notably by using pro-form substitution (but we also discussed constituency tests based on ellipsis, fragment answer and extraposition). Constituents may be discontinuous in Latin (as in other languages) but they do exist. They may not be easy to identify, but claiming that they do not exist without attempting to look for them is not a proper scientific attitude.

Università degli Studi di Milano-Bicocca
Dipartimento di Psicologia
carlo.cecchetto@unimib.it

Università degli Studi di Udine
Dipartimento di Studi Umanistici
renato.oniga@uniud.it

REFERENCES

Aelbrecht, L.
    2010 *The Syntactic Licensing of Ellipsis*, Amsterdam, Benjamins.

Brown, H.P. – Joseph, B.D. – Wallace, R.
    2009 *Questions and Answers*, in Baldi, P. – Cuzzolin, P. (eds.), *New Perspectives on Historical Latin Syntax*, Vol. 1, Berlin, Mouton de Gruyter, pp. 489-530.

Carnie, A.
2009   *Constituent Structure*, Oxford, Oxford University Press.

Cecchetto, C. – Oniga, R.
2002   *Consequences of the Analysis of Latin Infinitival Clauses for the Theory of Case and Control*, in «Lingue e linguaggio», 1, pp. 151-189.

Devine, A.M. – Stephens, L.D.
2006   *Latin Word Order: Structured Meaning and Information*. Oxford, Oxford University Press.

Chomsky, N.
1995   *The Minimalist Program*, Cambridge, MA, The MIT Press.

2012   *Problems of Projection*, in «Lingua», 18, pp. 1-35.

Cinque, G.
1999   *Adverbs and Functional Heads*, Oxford, Oxford University Press.

Danckaert, L.
2012   *Latin Embedded Clauses. The Left Periphery*, Amsterdam-Philadelphia, Benjamins.

Evans, N. – Levinson, S.C.
2009   *The Myth of Language Universals: Language Diversity and its Importance for Cognitive Science*, in «Behavioral and Brain Sciences», 32, pp. 429-492.

Giusti, G. – Oniga, R.
    2007   *Core and Periphery in the Latin Noun Phrase*, in Purnelle, G – Denooz, J. (eds.), *Order et cohérence en latin*, Genève, Droz, pp. 81-95.

Goldberg, L.
    2005   *Verb-stranding VP Ellipsis: A Cross-Linguistic Study*, Doctoral Dissertation, McGill University, Montréal.

Hankamer, J. – Sag, I.
    1976   *Deep and Surface Anaphora*, in «Linguistic Inquiry», 7.3, pp. 391-428.

Hofmann, J.B. – Szantyr, A.
    1965   *Lateinische Syntax und Stilistik*, München, Beck.

Holmberg, A.
    2007   *Null Subjects and Polarity Focus*, in «Studia Linguistica», 61, pp. 212-236.

Iovino, R.
    2011   *Word Order in the Latin DP: the Syntax of Demonstratives*, in Oniga, R. – Iovino, R. – Giusti, G. (eds.), *Formal Linguistics and the Teaching of Latin. Theoretical and Applied Perspectives in Comparative Grammar*, Newcastle-upon-Tyne, Cambridge Scholars Publishing, pp. 51-63.

Kühner, R. – Stegmann, C.
    1976   *Ausführliche Grammatik der lateinischen Sprache*, II.2, *Satzlehre*, 5. Auflage hrsg. v. A. Thierfelder, Hannover, Hahn.

Ledgeway, A.
    2012   *From Latin to Romance. Morphosyntactic Typology and Change*, Oxford, Oxford University Press.

Leo, F.

1896-1906 *Analecta Plautina de figuris sermonis*, I-III, Progr. Göttingen (= *Ausgewählte kleine Schriften*, hrsg. v. E. Fraenkel, Roma, Edizioni di Storia e Letteratura, 1960, pp. 71-184).

Marušič, F. – Žaucer, R.

2013 *A note on Sluicing and Island Repair*, in S. Franks (ed.) *Annual Workshop on Formal Approaches to Slavic Linguistics, The Third Indiana Meeting 2012*, Ann Arbor, Michigan Slavic Publications, pp. 176-189.

Merchant, J.

2001 *The Syntax of Silence: Sluicing, Islands and the Theory of Ellipsis*, Oxford, Oxford University Press.

2003 *Remarks on Stripping*, Ms. University of Chicago.

Merchant, J. – Simpson, A.

2012 *Sluicing: Cross-Linguistic Perspectives*, Oxford, Oxford University Press.

Oniga, R.

2014 *Latin: A Linguistic Introduction*, Oxford, Oxford University Press.

Pollock, J.-Y.

1989 *Verb Movement, Universal Grammar and the Structure of IP*, in «Linguistic Inquiry», 20.3, pp. 365-42.

Rizzi, L.

1997 *The Fine Structure of the Left Periphery*, in Haegeman, L. (ed.) *Elements of grammar*, Kluwer, Dordrecht, pp. 281-337.

Rosenbaum, P.

1967 *The Grammar of English Predicate Complement Constructions*, Cambridge, MA, MIT Press.

Ross, J.R.

1967 *Constraints on Variables in Syntax*, Ph.D. dissertation, MIT.

Zanuttini, R.

1997 *Negation and Clausal Structure: A Comparative Study of Romance Languages*, Oxford, Oxford University Press.