An argument for phrasal spell-out: Indefinites and interrogatives in Spanish*

Antonio Fábregas
IS, Universitetet i Tromsø

Abstract

In this article we will provide evidence in favour of Phrasal Spell Out (PSO), a procedure of lexical insertion where non-terminal nodes in a tree configuration can be targeted by spell-out. We will propose that the formal differences between two Spanish indefinite pronouns, *alguien* and *alguno*, can be captured if the morpheme *-ien* is analyzed as a lexical item which corresponds to a syntactic phrase; this phrase, crucially, is broken in the presence of a plural number projection. Independent properties of the internal syntactic structure of the interrogative make the lexical item *-ien* compatible with plural in that configuration.

1. Phrasal spell-out vs. head spell-out

Late insertion accounts work on the hypothesis that lexical insertion takes place after the syntactic structure has been built through merge. From this perspective, a particular lexical item spells out one of the nodes of the tree that comes as an effect of the application of syntactic rules. Spell-out is triggered by the existence of features which, in a perfect scenario, directly correspond to the entry of one and only one of the lexical items contained in the vocabulary of a particular language. In this perfect scenario, again, the properties of the tree should not differ from one language to the other, and grammatical variation should come as an effect of minimal differences in the feature content of heads and, more crucially, of the different lexical repertoire of each language (to the extent that the lexical repertoire connects with the phonological interface, thus leaving parametric variation to the PF branch of the grammar; cf. Chomsky 2008).

Several complications arise in this neat picture that make reality differ from the perfect scenario that we have just outlined. One of them is that, unless a lot of encyclopaedic information is introduced in the tree, some nodes would correspond to more than one lexical item. Consider, for example, the difference between *cat* and *dog* in English. These lexical items share all their grammatical properties (noun, count, animate...) and their differences are mainly encyclopaedic (because they denote different concepts), so they could be inserted

* I am grateful to Peter Svenonius, Gilliam Ramchand, Michal Starke, Éva Dékány and Luisa Martí for comments to previous versions of this article. All disclaimers apply.

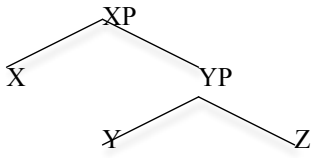
 Nordlyd 36.1, *Special Issue on Nanosyntax*, ed. Peter Svenonius, Gillian Ramchand, Michal Starke, and Knut Tarald Taraldsen, pp. 129–168. CASTL, Tromsø. http://www.ub.uit.no/baser/nordlyd/

in the same node of the tree. The difference between these lexical items and functional lexical items, such as agreement or case markers, is explored in Harley & Noyer (1998); we will not explore it in this article.

A second complication stems from the fact that some lexical items, even the functional ones, may correspond to more than one feature bundle in the syntax. This situation is known in morphology as syncretism, and can be illustrated, for example, by the English pronoun *me*. This pronoun is used to spell out first person singular both in the accusative case (*You saw me*), the dative (*You gave me a present*) and different oblique cases governed by instrumental, locative and directional prepositions (*with me, on me, to me*). These different cases presumably correspond to different syntactic configurations, and different information in the feature content of syntactic objects, but the lexical repertoire of English covers them all by the same vocabulary item. Mismatches of this kind between the syntactic information and the set of lexical items available in a language have been analysed in two different ways. The first one, commonly used in Distributed Morphology, is the Subset Principle (which states that a lexical item must have a (proper) subset of the features contained in the syntactic objec). Some syntactic features are possibly erased in this approach. The second one, proposed in Nanosyntax, is the Superset Principle (a lexical item must have all the features contained in the syntactic object and possibly some more; no syntactic feature can be erased or ignored by lexical insertion). In this paper we will assume the latter, the Superset Principle, and we refer to Caha (2009) for arguments in favour of it.

This article explores the third complication offered by the Late Insertion account. If lexical items are inserted in syntactic nodes, the immediate question that comes to mind is whether all nodes are equally fit for lexical insertion or, on the contrary, only some of them, with a particular property, allow for insertion of lexical material. The common view in Distributed Morphology and lexicalist theories is that only some of the nodes allow for lexical insertion: terminal nodes, also known as 'heads', are the only nodes where material can be inserted. In this way, inside the tree in (1), only the three terminal nodes can get material inserted, and, therefore, a minimal number of three lexical items would be used.[1]

(1)



[1] In this section we are only taking into account Late Insertion approaches. Notice, however, that in a Projectionist approach where no proper Late Insertion takes place it is necessarily the case that each lexical item corresponds to a terminal node. This follows from the fact that they are taken as atoms by the syntax and, as no internal syntactic structure is recognized in them, they can only be heads.

An alternative would be to consider all nodes equally fit for spell-out and allow for lexical insertion in any of the five nodes in (1), including the two maximal projections XP and YP. Insertion of lexical material inside a constituent would have the effect of checking all the features contained in the components of that constituent; thus, insertion in YP would satisfy the features of both Y and Z at the same time. Insertion in XP would satisfy the features of X, Y and Z.[2] This form of spell-out that allows for insertion in non-terminal nodes is known as Phrasal Spell-Out (PSO); by opposition, we will call the standard spell-out which is restricted to terminal nodes Head Spell Out (HSO).

PSO was first proposed, to our knowledge, in McCawley (1968) and was an important part of the generative semantics tradition, only that with the caveat that in this framework the tree was constituted of logical primitives which, sometimes, had no noticeable effect in the grammar of a language. Inside the modern Minimalist framework, it has been used in Weerman and Evers-Vermeul (2002) to give account of the interaction between person and case inside pronouns and by Neeleman and Szendröi (2007) to explain radical pro-drop patterns. PSO has been shown to be an interesting technical procedure that gives answer to some long-standing problems some of which we will explore in section five, but, to the best of my knowledge, no direct argument has been given in favour of it.

The paper is structured as follows. In the next section, we will briefly present the reasons why PSO is allowed under the present assumptions in Minimalism, although we will highlight the fact that PSO is not forced by the system and, therefore, requires empirical evidence. Section three explores one prediction of PSO theories which is not expected in a HSO approach; here we will present the behaviour of two indefinite pronouns in Spanish, *alguien*, 'someone', and *quién*, 'who', which will be used as evidence that the prediction made by PSO theories is borne out. Section four provides the analysis of the patterns described, and section five considers one advantage of PSO for morphological theories in general.

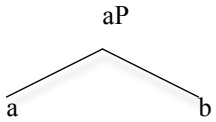
2. PSO is possible in a Minimalist Framework

Chomsky (2004) takes as his agenda to minimize the number of operations needed to account for the generative capacity of language. His proposal is that the tree is generated by using one irreducible operation, Merge. Merge is a recursive operation that takes two sets and makes a set out of them. The set thus created can be taken again and combined with another set in order to give as a result a larger set which contains the two previous sets as its members. Movement is reducible to Merge (Starke 2001, Chomsky 2004), and feature checking is integrated with Merge to the extent that the former triggers the latter.

[2] Here we assume the Inclusiveness Condition (Chomsky 2004), and, therefore, we assume the feature content of a maximal projection not to contain additional features not present in any of its components.

From this neat picture of Merge as the single basic operation in syntax, two consequences follow. The first one is that the tree is the representation of the hierarchical organization of sets, with the effect that movement is the situation where the same set is a subset of two different sets at the same time (from which the Multidominance approach to movement comes; cf. Gärtner 2002, Citko 2005, Svenonius 2005). The second consequence is directly relevant for PSO. As Merge combines two sets into a single set, it follows that terminal nodes are sets (singleton sets, more precisely) in the same way as maximal projections are sets. Otherwise, Merge would not be able to generate the first step of the tree, in which two terminal nodes are combined into a maximal projection (2). We conclude that there is no inherent difference in the nature of terminal nodes and non-terminal nodes.

(2)



One possible consequence of this approach is that both terminal and non-terminal nodes can undergo the same set of operations, among them being targeted by the insertion of lexical material. However, it is important to underline that this is not a necessary logical consequence of the approach, but rather an empirical question which has to be answered on empirical grounds, because the diagram in (2) allows for a configurational definition of terminal node. Notice that in the tree in (2), the diagram presented can geometrically define the nodes *a* and *b* as special in the sense that they do not dominate other nodes. It would be possible, then, that a rule of grammar makes reference to a particular subset of nodes, namely those defined inside the grammar as terminal. The question is, of course, if this kind of rule exists.

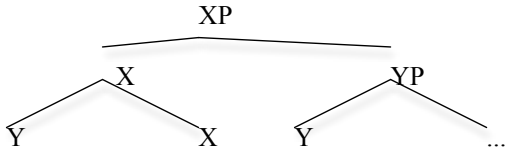
In this article we will argue for the idea that lexical insertion does not make a distinction between terminal and non-terminal nodes, and we will discuss an empirical phenomenon that shows how spell-out rules need to be able to make reference to non-terminal nodes, giving rise to PSO.

3. Different predictions of HSO and PSO

Theories are rejected or accepted by their success in giving account of data which were previously not understood. In this section we will argue for the adequacy of PSO by showing that one of its empirical predictions is borne out.

HSO restricts spell out only to terminal nodes, while PSO allows that any node is targeted by lexical insertion. These two theories differ with respect to the interaction that they predict between movement and lexical insertion. The geometric definition of terminal node (as a node that does not dominate other nodes) is not altered by movement operations. Even if head movement is allowed to take place in the syntax, a geometrically defined terminal node is still a terminal node before and after that operation.

(3)



In the configuration of (3), Y and X remain defined as terminal nodes, as they still do not dominate any other node. In the HSO approach, then, movement may change the ordering of particular lexical items linearizing to the left a constituent that was to the right, but no changes are expected with respect to the choice of the particular lexical items used in spell-out. These changes, in a system that restricts lexical insertion to terminal nodes, can only take place after non-syntactic operations are allowed in the system, such as fusion, impoverishment or the combination of both in Distributed Morphology (Halle & Marantz 1993), but they do not follow from movement precisely because movement does not change what counts as a terminal node.

However, when insertion in non-terminal nodes is allowed, we actually expect that movement or its absence will determine the particular lexical items that are selected for insertion. Non-terminal nodes are syntactic constituents, and the constituents that they contain may be different before and after a movement operation, be it phrasal or not.

For illustration, consider the following abstract situation. In (4) we have a tree structure which contains four heads. For the sake of exposition, let us assume that each one of these heads contains one and only one feature, so the capital letters stand both for the head and the feature carried by it.

(4) $[_{AP}A \quad [_{BP}B \quad [_{CP}C]]]$

Let us imagine that in this language the lexical repertoire contains a lexical item corresponding to the constituent in (4), as well as lexical items corresponding to the terminal nodes, as represented in (5).

(5) a. Lexical item 1 <--> {A{B{C}}}
b. Lexical item 2 <--> {A}
c. Lexical item 3 <--> {B}
d. Lexical item 4 <--> {C}

Here, PSO predicts that the Lexical Item 1 will be used to lexicalize (4). Now suppose that in the constituent (4), phrasal movement has taken place, changing the shape of the tree, as in (6).

(6) $[_{AP}\ [_{CP}\ C]$ A $[_{BP}\ B \quad [\sout{_{CP}\ C}]]]$

In this construction, the Lexical Item 1 cannot be used, as the phrasal configuration to which it is associated does not correspond to the diagram in (6). Here, the prediction is that the grammar would require the lexical items 2, 3 and 4 to lexicalize the same features. Thus, the structure in (4) would correspond to Lexical Item 1, while the structure in (6) would correspond to a different morphological make-up, Lexical Item 4 – Lexical Item 2 – Lexical item 3. Syntactic movement leads us to a different choice in the morphemes or roots used to lexicalize the same set of features. This situation, in which identical features are materialized with different morphemes, is generally classified in Morphology as allomorphy.

In consequence, if we do not allow post-syntactic operations as Distributed Morphology does, HSO and PSO make clearly different predictions with respect to the interaction between allomorphy and syntactic movement. While HSO does not predict any kind of interaction, PSO predicts that movement, to the extent that it changes the phrasal configuration to which (some) lexical items are associated, triggers allomorphy, as identical features would be spelled out by different lexical items. In the next section we will introduce an empirical phenomenon of Spanish that illustrates this kind of allomorphy and, as we will show, cannot be explained by a system that restricts lexical insertion to terminal nodes.

3.2. The behaviour of *alguien* in Spanish.

We will argue that the abstract situation presented in (4) and (6) takes place in natural languages. The empirical phenomenon that we will use to argue that spell-out is sensitive to movement comes from the contrast between two Spanish pronouns presented in (7).

(7) a. *alguien*, 'someone'
b. *alguno*, 'some'

There are several differences between the properties of these two pronouns. The first one is that *alguien* obligatorily refers to human entities; *alguno* can refer to humans, but does not require it. Thus, (8a) is necessarily interpreted as 'some person fell', but (8b) allows, additionally, a reading in which 'something fell'.

(8) a. Alguien cayó.
someone fell.3sg
'Someone fell'
b. Alguno cayó.
someone fell.3sg
'One fell'

The second difference is that *alguien* does not allow for plural forms, while *alguno* does. In other words, speakers are required to use *alguno* when they want to denote a group of humans (9).

(9) a. *Alguien-es cayeron.
someone-s fell.3pl
(Intended: 'Some people fell')
b. Alguno-s cayeron.
some-pl fell.3pl
'Some fell'

The third difference is that *alguien* does not allow partitive codas; that is, *alguien* cannot introduce a PP which denotes a group from which one specific individual is selected. *Alguno*, on the other hand, allows it.

(10) a. *Alguien de los niños cayó
someone of the children fell.3sg
(Intended: 'Some of the children fell')
b. Alguno de los niños cayó
someone of the children fell.3sg
'Some of the children fell'

Some authors have argued that *alguien* allows for some PP complements which seem to allow a partitive interpretation; let us see why this position is wrong. Among the cases where partitive codas have been claimed to appear with *alguien* we find the phrases in (11).

(11) a. alguien de los aquí presentes
someone of the here present.pl
'someone from those present here'
b. alguien del ejército
someone from-the army
'someone from the army'

Considerable amount of empirical evidence shows that these PPs are not partitive codas, as they do not have their formal or semantic properties. Notice, to begin with, that the English gloss requires the preposition *from*, not *of*, which is generally used in partitive and pseudopartitive constructions in this language (*two of my friends; a group of soldiers*). Gutiérrez Rodríguez (2008) mentions, as further evidence that the examples in (11) are not partitive constructions, that these PPs allow the preposition *entre*, 'between, among', which true partitive complements reject (12); they marginally allow right dislocations, while partitive complements don't (13); they can contain quantifiers, while partitive codas reject them (14); they can take collective nouns –denoting groups, but formally singular, such as *army*, *family* or *staff*-, while partitive codas don't (15).

(12) a. alguien de entre el ejército
someone from between the army

'someone from among the army'
b. *dos de entre los amigos
two of between the friends
(Intended: 'two of the friends')

(13) a. Alguien vino de los aquí presentes.
someone came from those here present.pl
b. *Dos vinieron de los amigos.
two came.pl of the friends

(14) a. alguien de todos los presentes
someone from all those present
b. *dos de todos los amigos
two of all the friends

(15) a. alguien de la clase
someone from the class
b. *dos de la clase
two of the class

The phrases in (16) show that *alguno* takes real partitive codas, as it rejects phrases with the properties illustrated in the previous series of examples.

(16) a. *alguno de entre ellos
some of between them
b. *Alguno vino de ellos
some came.sg of them
c. *alguno de todos ellos
some of all them
d. *alguno de la clase
some of the class

Alguno and *alguien* are equal, though, in that they cannot combine with nouns, as illustrated in (17). This generalization can be refined, though, in the case of *alguno*. It seems that it is the presence of the desinence *-o* which blocks the presence of an overt noun in this context (see the contrast between 17b and 17c). It is a general property of *un-*, as the indefinite article, and the quantifiers that contain it, that the desinence *-o* in the singular form and an overt noun are in complementary distribution. The incompatibility disappears, though, in the plural and feminine forms (18).

(17) a. alguien (*niño)
someone child
b. algun-o (*niño)
some boy
c. algún *(niño)
some boy

(18) a. algun-a niña
some-fem girl

b. algun-os niños / algun-as niñas
some-masc.pl boys / some-fem.pl girls

The different behaviour of the desinence in the masculine singular and in the rest of the inflectional forms is puzzling, but we will not discuss it in this article. Here we will concentrate on the difference between the form *alguno* and *alguien*.

3.3. Why this behaviour cannot be captured by a HSO theory

It is interesting to consider how the different behaviour of *alguien* can be captured in a system where *-ien* (alternatively, *alguien*) corresponds to a head. As is generally the case in approaches where lexical items correspond to heads, the properties that the item exhibits in the syntax depend on the content of the bundle of features contained in a single head. The fact that *alguien* cannot have a plural form suggests, in that case, that the feature content of the lexical item is not compatible with a plural feature. By stipulation, this kind of approach could lexically specify the item as being [singular]. However, this stipulation would not be enough to prevent *alguien* from combining with a Number Phrase (Ritter 1991) specified as plural, as nouns like *car, man* or *doctor* can also be interpreted as singular in the absence of the plural morpheme, and though they are able to combine with a Number Phrase which contains plural information. This approach, thus, would need to treat *alguien* as a singularia tantum, that is, a noun which is lexically frozen as a singular and cannot combine with plural information in any case. This groups *alguien* with a very reduced set of nouns in Spanish which do not take plurals, such as *hambre*, 'hunger', *frío*, 'coldness', or *sed*, 'thirst'.

One problem is, however, that the reduced set of Spanish singularia tantum are characterized by some specific properties which *alguien* does not share with them. Singularia tantum in Spanish are systematically mass nouns, while *alguien* is clearly a count pronoun (see below for formal evidence of this), and they almost always refer to psychological states. At this point, treating *alguien* as a singularia tantum would be a pure stipulation, as stipulative as to impose that it cannot be combined with plural information.

However, even if there was a better way of imposing that *alguien* represents a head which cannot combine with a plural head, several problems would remain. The first of them is that such stipulation would not explain why *alguien* cannot take partitive complements. The pronouns in a partitive construction can be in the singular form, as shown in (19), and yet *alguien* cannot occupy that position in the structure, as noticed in (10).

(19) uno de los niños
one of the children

The fact that the pronoun *alguien* both rejects the plural morpheme and the partitive construction, while the pronoun *alguno* accepts both of them, does not seem to be accidental, suggesting a connection between the two properties which

cannot be capture by merely stipulating the incompatibility between two syntactic heads. Additionally, notice that stipulating that this pronoun is marked as singular in the lexicon presupposes that [singular] is a feature, an assumption that we will challenge in this article.

Another, perhaps more crucial, reason not to stipulate that *-ien* cannot combine with the plural head, is that the same lexical item is compatible with this information in other contexts. Consider the interrogative *quién*, 'who'. This word can be transparently decomposed in two morphemes, *qu-*, shared by other question words in Spanish, and *-ien*. Evidence that this *-ien* is the same item that we find in *alguien* comes from the following facts: a) both words only behave as pronouns and can never combine with overt nouns (cf. 20a) and b) both words refer only to human entities (cf. 20b).

(20) a. ¿Quién (*niño) vino?
who (child) came.sg?
b. ¿Quién cayó?
who fell.sg? ('what is the x, x a human, such that x fell?)

Additionally, the two lexical items share the same historical origin. Once we have established that there are no reasons to think that *-ien* is a different lexical item in the interrogative, consider the data in (21). Here, the interrogative clearly combines with a plural morpheme, which triggers agreement in plural with the verb.

(21) ¿Quién-es vinieron?
who-pl. came.pl?
'which people came?'

Stipulating that *-ien* cannot combine with the plural head has, thus, the problem that the same lexical item seems to be able to combine with plural information in other contexts. Therefore, an analysis where *–ien* corresponds to a single head will be forced to stipulate that this head cannot combine with the plural in some cases. The alternative of treating *alguien* and *quién* as single morphemes (that is, not segmenting *-ien*), and stating that only the first is incompatible with the plural information is untenable as it misses the generalization that in both pronouns the form *-ien* imposes animate readings not forced in its absence.

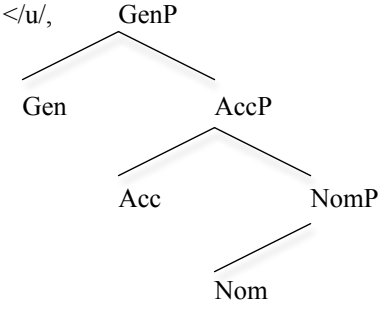
4. The analysis

4.1. The rules of the game: how to perform Lexical Insertion with PSO

The PSO has as its basic proposal that (some) lexical items are associated not to single heads, but to syntactic configurations. Thus, the shape of a lexical item would be, maximally, as in (22). The lexical item is a relation between a phonological exponent and a syntactic configuration; in case the item corresponds

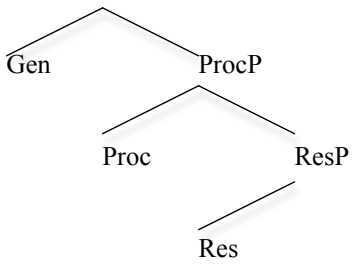
to what has traditionally been called 'a lexical category', to this information, semantic concepts are added. Otherwise, the lexical item does not contain conceptual information, as the meaning would be entirely carried by the features inside the syntactic configuration.

(22) a. </u/, GenP >



[adapted from Caha 2009: 53]

b. </ə□raiv/, GenP, CONCEPT>



[adapted from Ramchand 2008]

The lexical entry in (22a) represents the genitive case in Modern Greek, in a word such as *vun-u*, 'mountain', genitive corresponding to the nominative form *vun-o*. In this lexical entry it is not necessary to represent any concept, as the semantic information required is already contained in the features to which the phonological exponent is associated. In contrast, the lexical entry in (22b) corresponds to the verb *arrive* in English, and needs to add conceptual information which is not contained in the formal features to which the item is associated.

Given these entries, let us see how the procedure of lexical insertion is performed. The first principle which drives lexical insertion is that every syntactic feature inside the structure needs to be identified by lexical insertion. In other words, features which are represented in the syntax cannot be ignored by lexical insertion, which means that Impoverishment (the procedure that erases in PF some features from the syntactic representation) does not exist in this system.

We will call this principle the Exhaustive Lexicalization Principle (Fábregas 2007).

(23) Exhaustive Lexicalization Principle.
Every feature inside the syntactic representation must be identified by lexical insertion.

The Exhaustive Lexicalization Principle has an immediate consequence for the treatment of syncretism, the situation where a single lexical item lexicalizes two or more different sets of features. Given that syntactic features cannot be ignored by lexical insertion, any mismatch between the lexical entries and the syntactic structures needs to be solved by inserting a lexical item that has more features, and not less, than the syntactic configuration. This principle is called The Superset Principle and is due to Michal Starke. In (24) we reproduce the definition given in Caha (2009: 55).

(24) The Superset Principle.
A phonological exponent is inserted into a node if its lexical entry has a (sub-)constituent which is identical to the node (ignoring traces).

Let us illustrate this principle from previous work by Caha (2009). Consider, for example, the case of the nominative, the accusative and the genitive of the root *vun*, 'mountain' in Modern Greek (25).

(25) Nom: *vun-o*
Acc: *vun-o*
Gen: *vun-u*

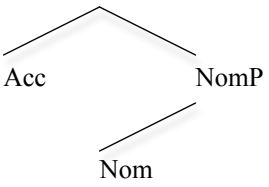
The nominative and the accusative are syncretic in this paradigm. In a system where syntactic features can be ignored by lexical insertion, the syncretism is normally performed by assuming that the nominative case is used in the accusative form after the accusative matrix of features has been impoverished, becoming identical to nominative. This situation where the exponent associated to the smaller set of features has been known as the Subset Principle (cf. Halle 1997). However, this description encounters a problem when other words are considered in Modern Greek. (26), the partial paradigm for *anthropos*, 'man', shows that, when syncretism does not take place, *-o* is the exponent for accusative, not nominative.

(26) Nom: *anthrop-os*
Acc: *anthrop-o*
Gen: *anthrop-u*

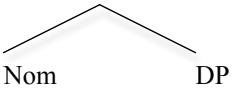
This shows that it is misleading to treat the syncretism in (25) as nominative extending to accusative. The form that gets extended is the accusative, which, by hypothesis, has more features than the nominative (see also Caha 2009 for

evidence of this, taken from Blake's (2000) hierarchy). The empirical result is predicted by the Superset Principle. The entry for the accusative is the one in (27a); if the root *vun* lacks a particular exponent for its nominative case marking, the accusative is inserted in (27b), as the tree is identical to a subconstituent of its lexical entry. This situation where a lexical item is used to spell out a smaller configuration is known as Shrinking.

(27) a. </o/, AccP >

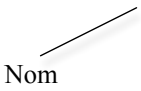


b.



What happens with *anthropos*, then? In this noun stem there is a designated lexical exponent for nominative, *-os*, with the entry in (28).

(28) </os/, NomP>

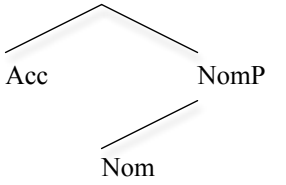


An Elsewhere Condition is necessary, as in all accounts that deal with competition between lexical items, in order to guarantee that *-os*, and not *-o*, will always be used in the nominative of *anthropos*. As in other cases, the Elsewhere Condition states that, given two lexical items competing in the same context, the one that will be chosen is the one that has fewer unmatched features. Given the Superset Principle, the one chosen in this context will be the one with the smaller number of features, as it will be this one which has with the most restricted distribution. In a nominative context, then, *-os* is preferred to *-o* when both items are available, because the first can only be used in a nominative context, while the second can be used both in nominative and accusative contexts.

Empirical research made within frameworks where lexical items are allowed to correspond to phrases has shown that the situations in which lexical items are allowed to shrink crucially imply always the lowest projections. This is known as the Anchor Condition (Abels & Muriungi 2008, Caha 2009), which specifies that the lowest feature in a lexical entry must always be matched with the syntactic structure. In the case of *vun* discussed before, notice that our procedure must allow the accusative marker to shrink and express only the nominative layer; however, allowing it to shrink to express only the accusative layer in any case

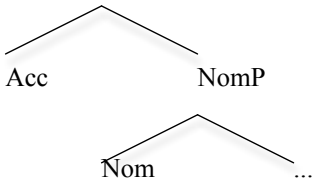
would predict that a word such as *anthropos* could get the form **anthrop-o-os* in the accusative, against the empirical facts. Consider why. (29a) reproduces the lexical entry for the accusative marker; (29b), that of the nominative marker. In a tree such as the one in (29c), corresponding to the syntactic configuration for accusative case, the Anchor Condition guarantees that /os/ would not be inserted in the nominative layer, because that would force the accusative /o/ to shrink to Acc, which is not its lowest node in the lexical entry. Given that the Accusative Layer needs to be lexicalized by the Exhaustive Lexicalization Principle, the only possibility to lexicalize (29c) is by matching the accusative exponent with both syntactic heads.

(29) a. </o/, AccP >



b. </os/, NomP>

c. </u/, AccP >



Notice that throughout this presentation we have consistently ignored the complement of the lowest head in the lexical entry. This comes from the assumption that the complement of the lowest head (DP in 29c) must have been displaced before the lexical insertion takes place. Under this situation, the element occupying that position for the time that lexical insertion takes place is a lower copy, which, following general assumptions (cf. Fox 2000, among many others), is ignored by lexical insertion. We will suggest in section 4.5. that the presence of a lower copy in the lexical entry of an item has direct consequences for shrinking, although the implications of our suggestion are not developped in this paper.

Given the procedure of lexical insertion assumed in this paper, and which we have just presented, let us see how the analysis of the pattern described in section 3 is performed.

4.2. The Fseq

The first step is to identify the syntactic constituents involved in the lexical items *alguien* and *alguno* which we are discussing in this paper. We assume a cartographic approach to syntax where there is a fixed fine-grained functional sequence (Fseq). Inside the Fseq, different areas can be identified; inside each one of these areas, heads that share crucial aspects of their behaviour and of their semantic and formal contribution to an utterance are organized in an ordered fashion. The motivation to place a particular head in a specific position inside the Fseq comes from the principles of semantic compositionality, such as the higher heads operating on the information provided by the lower heads. Due to this, the default assumption with respect to the Fseq is that it is universal, as proposing different orderings in different languages would lead to LF-parameters, whose status is dubious in present research.

We follow Cinque (2005) in the assumption that inside the nominal domain four distinct areas can be identified: the area for the determiner, the area for number, the area for adjectives and the area for nouns. This author shows that the assumption that these areas are ordered as Det > Num > Adjective > Noun is able to derive the typology of word orders inside this domain, and therefore we will assume this ordering.

The Number area must be further decomposed, though. Inside the number area, we find at least two different types of heads. The first kind of head defines grammatical number, and in a language like Spanish includes singular and plural as its values; other languages contain further specifications, such as dual, trial and paucal. The second kind of head are quantifiers, that is, operators that map singular entities or groups into quantities, and that, unlike number, give rise to scope phenomena. The ordering between these two types of heads is generally assumed to be Quantifiers > Number. Data coming from languages where quantifiers do not exhibit morphological agreement with nouns show that the lexical items which spell out these features systematically appear to the left of those that spell out number.

(30) a. many actor-s
b. (Norwegian) mange skuespiller-e
many actor-pl

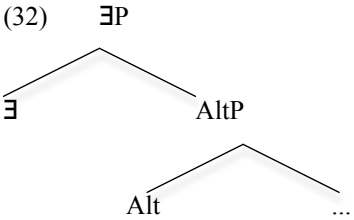
In a more fine-grained decomposition among the different quantifiers, several morphological generalizations can be made which throw some light on the ordering between the heads in this level. Wh-operators are systematically lower than indefinite quantifiers; this is shown by morphological evidence, as indefinites are frequently constructed on top of wh-words, as shown in (31).

(31) a. where – somewhere

b. (Hungarian) ki – valaki
who – someone
c. (Russian) kto – koe-kto
who - someone
d. (Latin) qui – qui-dam
who – someone
e. (Khmer) noa – neaq-noa
who – someone

In analysing wh-words it is crucial, however, to be able to discriminate between the information which is provided inside the nominal domain and that which is due to the CP projections that add interrogative force. Languages with systematic wh-in-situ, such as Japanese, show that the interrogative word (once we have factored out the interrogative illocutionary force) denotes a set of alternatives. In these languages, there are no interrogative words, but indeterminate pronouns, and the interrogative interpretation arises when a designated interrogative operator is merged outside the nominal domain, marking the scope of the interrogation. In these languages, the indeterminate pronoun denotes a set of alternatives from which the answer to the question must be picked up (labeled as an 'indeterminate pronoun' in these languages); the set of alternatives denoted by the indeterminate pronoun needs to be long-distance bound by a relevant operator, which can be an interrogative operator or any other kind of quantifier (Kratzer and Shimoyama 2002). Assuming, as these languages suggest, that a wh-word is constructed always over a head that denotes a set of alternatives, the decomposition of the relevant part of the quantifier area would be as in (32).

(32)

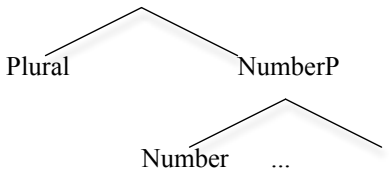


Before moving to the CP domain, then, the wh-word denotes a set of alternatives (not an individual); over this layer, an existential operator can be merged, giving as a result an indefinite pronoun. This existential operator binds AltP. In this way, AltP without the existential operator denotes a set of alternatives, taken from the possible objects that comply with a particular description. With the existential operator, it denotes that there exists at least one entity which complies with the description over which the set of alternatives was constructed. This implementation straightforwardly explains why indefinites frequently contain wh-morphology, but lack the semantics associated to wh-words when they occur alone. As the AltP needs to be bound by an operator, introducing the existential area over it satisfies this condition; in the absence of the existential layer, AltP

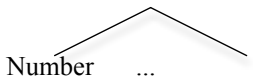
still needs to find an operator, and finds it in the CP domain in the form of the interrogative operator.

As for the number area, which we assume to be lower than the quantifier area, we propose to decompose it in at least two heads, following suggestions by Michal Starke. The first node denotes a number value, while a second head that dominates this one denotes specifically plural. The combination of both sets, as in (33a), denotes plural number, while the set formed by Number in the absence of Plural denotes singular number (33b). This explains that a noun phrase without any specific number marking can combine with the plural feature without incurring in semantic contradiction. If the noun phrase without the plural feature contained a feature [singular], combination with the plural feature, manifested as *-(e)s* in Spanish, would give rise to a semantic contradiction, as the same object would contain information of singular and plural. In contrast, if we take singular number to be the absence of the plural feature in a context where there is number information, the contradiction evaporates. Notice also that, in generic contexts, nouns in 'singular' generally denote groups of entities (*A first rate student is always full of questions*), which further supports the idea that 'singular' does not exist as a designated head or a feature.

(33) a.



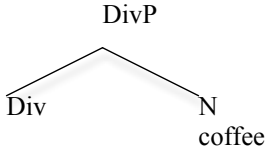
b.



Notice that in an analysis such as this one, where 'singular' does not exist as a feature, it is impossible to lexically specify a noun or a pronoun as being restricted to singular contexts in its lexical entry. In consequence, the incompatibility of *alguien* with a plural feature has to be explained in another way.

Let us consider now the fine-grained decomposition of the noun phrase. It is generally agreed (Borer 2005) that the distinction between count and mass nouns, which is crucial for the expression of number, is represented by a specific syntactic head, the Divisor. The divisor takes an entity which denotes 'stuff' and turns it into a package of stuff, such as that it can be enumerated, as in (34). In this simple example, the node NP denotes a substance which we normally refer to as 'coffee'; the presence of the divisor takes that substance and packs it into count units.

(34)



In languages such as Spanish, it has been claimed that the Divisor is lexically spelled out as the desinence (Picallo 2006, Alexiadou & Gengel to appear), which sometimes attaches to the noun and to some pronouns. The desinence also contains information about the lexical class to which the NP belongs (for which Borer uses the term Classifier to refer to the Divisor, merging in the same head these two properties). Thus, we assume that the desinence contains both information about the class to which the noun belongs and about the count-mass properties of the noun (see also Fábregas & Pérez 2008 for independent evidence of this).

Animacy seems to be, cross-linguistically, a property codified in noun phrases which has clear implications for agreement and other syntactic processes (Corbett 1991). The interaction between animacy and agreement shows that this has to be represented as a formal feature in syntax, and not as part of the conceptual information added at the interfaces. We propose that the feature must be placed on top of DivP. The motivation for placing the feature in this position is that animate nouns must contain gender, but nouns with gender do not need to be animate. As shown in (35a), a noun such as *table* contains a desinence, and gender information, even though there is no animacy involved. Additionally, when a noun is animate, the animacy needs to access the information about noun class and gender in order to determine if the noun denotes the male or the female member of the animate entity denoted (35b,c). The fact that animacy requires gender, but not the other way round, combined with the evidence that shows that animacy must take into account the gender information motivates the Animacy feature being higher in the structure than DivP.

(35) a. la mes-a alt-a
the.fem table-fem high-fem
b. el niñ-o
the.masc child-masc
'the boy'
c. la niñ-a
the.fem child-fem
'the girl'

On the assumption that every feature introduced in the syntactic derivation must be interpreted, it follows that the animacy feature will only be present in those structures which must be interpreted as animate; for our purposes, this means that *alguien*, which cannot be interpreted as non-animate, contains the animacy feature. In contrast, *alguno,* which allows, but does not force, an animate interpretation, does not contain this feature. If it contained it, we would be forced to assume that in some cases, even though the feature is present, it is not interpreted, which goes against the assumption that we just presented.

There is a further property that defines nouns as opposed to other categories. Cross-linguistically, nouns are differentiated from other categories in that they introduce an index of identity (Baker 2003). The index of identity is used to make

judgements about whether two different descriptions refer to the same entity or to different entities, and, as discussed in Baker (op.cit.), these indexes must take into account the properties denoted by the base NP. The two sentences in (36) do not entail each other, as it may be the case that the same person was counted as three passengers, because he or she took part in three separate trips.

(36) a. The number of people that used SAS was 300,000 last year.
b. The number of passengers that used SAS was 300,000 last year.

It can be shown that the identity feature also interacts with the noun class, as the sentences in (37) do not entail each other: the individuals denoted by the feminine form may be included in the denotation of the masculine form.

(37) a. El número de pasajeros que usó SAS fue 300 el año pasado.
The number of passengers.masc that used SAS was 300 the year last
b. El número de pasajeras que usó SAS fue 300 el año pasado.
The number of passengers.fem that used SAS was 300 the year last

Thus, the identity index needs to have access to the gender information, which is contained in the DivisorP. The identity feature needs to differentiate between animate and non animate entities. We propose, then, that this feature is introduced above both DivP and AnimacyP. We believe that this is necessarily imposed by the assumption that meaning is compositional: assuming that a head needs to have in its domain all the information that it requires to satisfy its semantic contribution to the sentence, it follows from the fact that the index of identity needs to access gender information that the projection introducing it must be higher than the one that introduces gender.

Following Baker (2002), we assume that the feature is introduced by little n, which defines the whole projection as a fully independent noun phrase. Notice that, under this assumption, the identity index must be higher than Animacy and Divisor. The reason for this is that in a language such as Spanish, nouns need to have gender information, but adjectives, which are not nouns, can also have this information (remember (35a)). In other words, being a noun presupposes having gender information, but having gender information does not imply being a noun. Also, notice that some adjectives and adjectival participles are restricted to animate entities (such as *intelligent, worried*, etc.), which shows that it is possible to contain this feature without containing the nP. All these considerations motivate nP being higher than AnimacyP and DivP. A similar reasoning can be made with respect to certain verbs, which can only assign theta roles to animate entities (*think, know*...).

Notice that the data force us to conclude that nP in Spanish can be lexicalized by a zero morph, ∅. We have already shown that desinences do not lexicalize nP, and the morphological evidence in adjectives and nouns shows us that there are no additional lexical items.

The decomposition of the Noun Phrase area is, thus, as presented in (38), where the adjective area is not introduced, as it will not be relevant in our analysis.

(37) [∃P [AltP [PluralP [NumP [nP [AnimP [DivP [N]]]]]]]]

4.3. Why *alguien* rejects the plural

At this point we will determine how the lexical items that compose *alguien* and *alguno* are mapped into the structure that we have identified. In order to do this, we will take into account the fine-grained semantic analysis in Martí (2008).

The first observation that can be made is that *alg-*, the first component shared both by *alguien* and *alguno*, is a quantifier.

Verbs that select a quantifier as their complements, such as *weigh* or *measure* in their stative readings, can take *algo*, on a par with QPs; DPs and bare NPs are rejected (38).

(38) a. El bebé pesa tres kilos.
the baby weighs three kilos
b. #El bebé pesa los (tres) kilos
the baby weighs the (three) kilos
c. #El bebé pesa kilos.
the baby weighs kilos
d. Juan pesa algo
Juan weights some
'Juan has some weight'

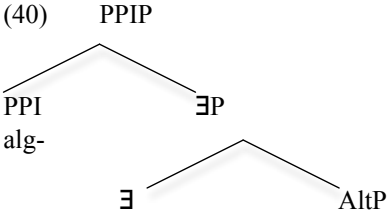
The meaning of (38d) is 'there is an undetermined quantity that corresponds to Juan's weight'; notice that no partitive presupposition is needed here, in the sense that it is not necessary to understand that the something is a part of a bigger set. The question at this point is what kind of quantifier is lexicalized by this item. The sentences in (39) show that *algo* and the pronouns that contain *alg-* systematically behave as polarity items. (39a) is interpreted as 'there was something in particular that she did not say', not as 'she did not say anything'; (39b) and (39c) mean 'there was someone that did not come'.

(39) a. No dijo algo.
not said.3sg something
b. No vino alguien.
not came.3sg someone
c. No vino alguno.
not came.3sg someone

Martí (2008) convincingly argues that the layer that implies movement above negation is hierarchically higher, inside the Fseq, than the layer that introduces existential quantification and presupposition. As this property is contained in all the words that contain *alg-*, it seems plausible to think that in Spanish the relevant

layer of the Fseq is lexicalized by this morpheme. From here it follows that the position of *alg-* inside the Fseq is the one in (40), where PPI stands as an abbreviation for 'positive polarity item', that is, the layer that forces the structure to scope out of negation.[3]

(40)

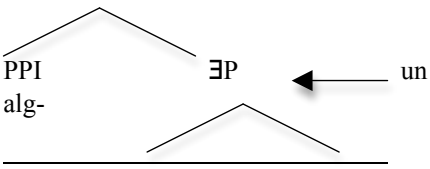


The lexical item *-un*, contained inside the indefinite *alg-un-o*, is at least homophonous to the indefinite article in Spanish. We will argue that the two items are not homophonous, but the same element. This morpheme and the indefinite article also share a further property, which is that in masculine singular forms, the desinence *-o* is compulsory if there is no overt noun and impossible otherwise (41). We take this as evidence that we do not have two morphemes that happen to sound the same, but actually two instances of the same morpheme.

(41) a. un(*-o) niño.
b. alg-un(*-o) niño.
c. un*(-o).
b. alg-un*(-o).

We assume an analysis of the indefinite article in Spanish in which it stands for an existential quantifier (see Elbourne 2002 on how to analyse donkey anaphoras and similar cases without treating the indefinite as a variable, against Kamp 1981 and Heim 1982). Although it is a highly complex problem (cf. Gutiérrez Rexach 2003: 21-137), we will assume here this analysis. Notice that associating *-un* to the existential layer establishes a nice correlation between morphology and semantics, as it complies with the semantic analysis in Martí (2008).

(42)

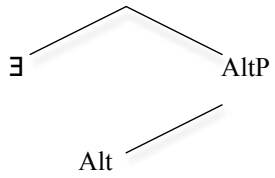


[3] Notice, however, that in Martí's (2008) account the plural indefinite *unos* is also analyzed as a positive polarity item, which to the extent of our knowledge is borne ut by the data. It seems, then, that somehow the PPI layer can be filled by procedures that do not imply inserting *alg-*. Pending further research on why *unos* can fill the PPI layer, the fact that all words that contain *alg-* behave as PPI shows that this morpheme is associated to that layer inside the structure.

∃ AltP

When both the existential and AltP are introduced in the structure, the whole constituent is phrasally spelled out as *un-*. We propose that, when only AltP is introduced in the structure, there is another lexical item which is specialized to spell out this single head, *qu-*. Thus, there are two competing lexical entries in this domain:

(43) a. </un/, ∃P >



b. </k-/, Alt>

Given the Elsewhere Condition introduced in 4.1., in the presence of the tree in (42), *un* is introduced, while if the existential is lacking, *qu-* is introduced, as it contain less unmatched features.

Let us consider now which part of the structure is associated with the desinence *-o*. Remember that in this article we will restrict the discussion to the masculine singular form *alguno*, and therefore we will not address the independently interesting question of how to account for the interaction between the different gender values and number.

The first generalization that has already been made is that the presence of the desinence *-o* in the singular is not compatible with an overt noun (remember (41)). This property is captured immediately if the desinence spells out a constituent which contains the lowest head in the nominal domain, NP, given the Anchor Condition. If overt nouns are inserted in this position, the fact that they do not co-occur with the desinence in this context is explained if the desinence is spelling out this NP. Are there other heads that *-o* can spell out? Following Picallo (2006) and Alexiadou & Gengel (to appear), the desinence in Spanish needs to be able to also spell out the DivP. The evidence that these authors provide involve the trivial fact that the desinence varies with the noun class of the noun, but, more crucially, that the desinence cannot license noun ellipsis when the noun is not count, and, therefore, the DivP is missing (44).

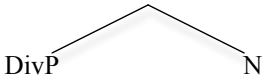
(44) a. Aquí hay café árabe (*y un-o colombiano)
here there-is coffee arabic and one colombian
b. Aquí hay chicos árabes y un-o colombiano.
here there-are boys arabic and one colombian

Evidence that it is the desinence, and not the presence of the indefinite article, that causes the incompatibility with the mass noun comes from the fact that the indefinite article, as expected from an existential quantifier, is compatible with mass nouns (45).

(45) Aquí hay un agua fresca y pura.
here there.is a water fresh and pure

This set of cases shows that the desinence needs to spell out a constituent that goes from DivP to NP, as represented in (46).

(46) < /o/, DivP >



Remember that in *alguno*, as opposed to *alguien*, the animate interpretation is not compulsory. For this reason, we propose that the animacy feature is not lexicalized by the desinence *-o*. By transitivity, everything on top of the animacy feature cannot be lexicalized by the desinence *-o*. Notice, in any case, that the desinence cannot be lexicalizing the little n. The reason for this is that in contexts such as those in (46), if the desinence lexicalized n, we would expect the word to be compulsorily a noun, which goes against the facts.

(47) roj-o
red-desinence
'red'

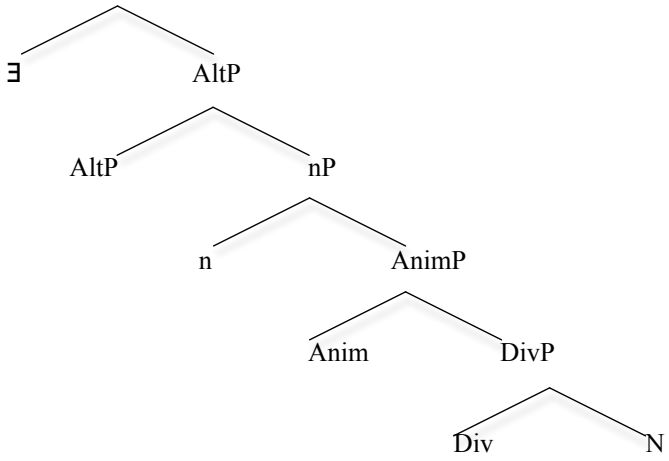
Instead, in our analysis, this desinence is an agreement marker that checks the values for the noun class without imposing any grammatical category to the word. In our proposed structure, notice that nP is built on top of the DivP. This guarantees that a word which is a noun needs to be assigned to a class noun, but at the same time that it is possible to be associated to a gender without being a noun, to the extent that DivP does not contain the nP in its domain.

Let us consider now what set of features *-ien* is associated with. By comparing it with the structure introduced by *alguno*, we can infer that *-ien* lexicalizes the existential, AltP, NP and the Divisor. The difference between *alguien* and *alguno* is that *alguien* only refers to humans, which implies that it also lexicalises the Animacy feature. However, all these properties can be motivated. (48a) shows that *alguien* is an indefinite pronoun; notice that it cannot be the contrastive focus in a sentence. (48b) shows that it is a count noun, as it cannot be the subject of a symmetric predicate, which requires to take a collectivity of individuals or a mass as its argument. Remember that earlier we showed that *alguien* does not combine with overt nouns, which shows that it also lexicalizes NP, and that it is restricted to humans, which shows that it lexicalizes AnimacyP.

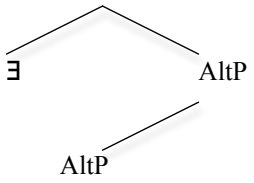
(48) a. *Ha venido ALGUIEN.
has come SOMEONE
'There has been someone who has come'.
b. *Alguien se amontona ahí.
someone SE piles up there

This motivates that *-ien* is associated with a lexical entry like the one in (49a), where the constituent lexicalized by it includes what *un-* (49b) and the desinence *–o* (49c) lexicalize, plus the animacy feature. The item *alg-* lexicalizes the same feature in both cases (49d). In these entries we do not yet represent the number area, as the interaction with it will be discussed in detail in the following pages.

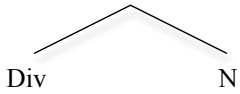
(49)a. </ien/, ∃P >



b. </un/, ∃P >



c. </o/, DivP >



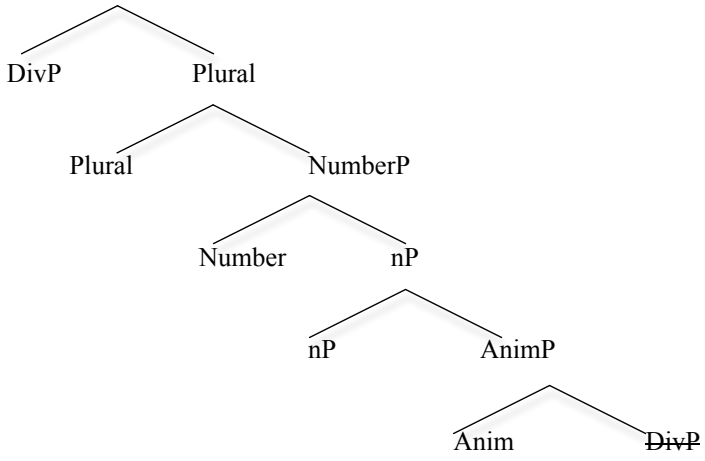
d. </alg/, PPI>

At this point we can start explaining why *alguien* is not compatible with the plural feature. Notice that the constituent that *-ien* lexicalizes crosses from the noun area to the area where indefinites are built. In the middle of this area we would have the information about number. More crucially, in an indefinite noun phrase, the denotation of the phrase crucially takes into account the interaction between the DivP and plural. In the general case, a noun in the plural form must be count, as the plural semantics is constructed by turning single atoms, which correspond to the entities packaged by the DivP (Borer 2005), into a group (50a). Remember that the singular does not impose the reading in which there is more than one object (for example, in generic contexts, (50b)), which motivates in our account the fact that singular does not exist as a feature. Notice also that mass nouns are generally in the singular, which shows that NumberP and DivP do not need to interact. In contrast, PluralP and DivP do.

(50) a. boys ⇒ a group which contains more than one boy
b. A boy is always taller than a girl.
c. water, sand, flour, air, poison...

We propose to capture this interaction between PluralP and DivP by proposing that PluralP attracts DivP to its specifier to configurationally denote a group formed with the atoms denoted by DivP, as in (51). Notice that this movement is backed by morphological evidence, as in Spanish the plural morpheme is a suffix attached to the noun stem, including the desinence (51b).

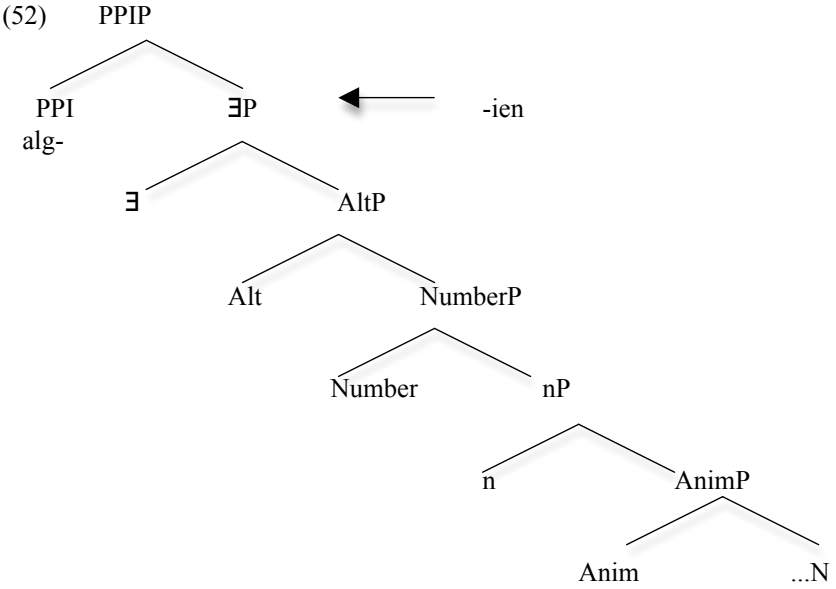
(51) a. PluralP



b. [$_{PlP}$[$_{DivP}$ perr-o] s]
dog-masc pl

At this point, the reason why *alguien* rejects the plural is straightforward. In the tree in (52a), the syntactic constituent that *-ien* requires has been constructed. Notice that we have to assume that *alguien* also lexicalizes NumberP, which actually follows from the data, to the extent that *alguien* triggers number agreement with adjectives and verbs (52b). *Alg-* is introduced in PPI, and *-ien* lexicalizes the whole ∃P.

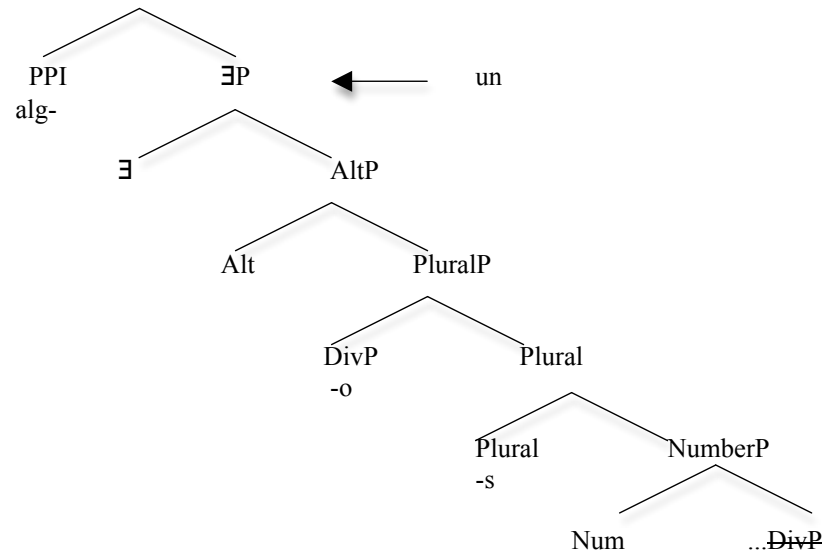
(52)



Now, consider what happens if PluralP is introduced. In this situation, DivP is attracted to its specifier (53a), breaking the constituent required by *-ien*. The Superset Principle would allow *-ien* to shrink to lexicalize only DivP, but remember that, by the Elsewhere Principle, this will be blocked by any lexical item which has fewer unmatched features and is therefore more specific. Indeed, in Spanish, there is such a lexical item, the desinence *–o*. This makes it impossible to use *-ien* in this context. In case there is no lexical item able to lexicalize AnimacyP on its own, this derivation would be impossible in any situation, as the feature would not be lexicalized, but this is independent of our analysis, as the movement itself makes it impossible that *-ien* is used in the structure.

Once the desinence is used to lexicalize the lower part of the tree and *–ien* cannot be inserted in this configuration, *-un* needs to be inserted to cover the higher part of the tree. This gives rise to the morphological decomposition lexicalization shown in (53b).

(53) a. PPIP



b. [$_{PPIP}$ alg [$_{\exists P+AltP}$ un [$_{PlP}$ [$_{DivP}$-o] s [$_{NumP+nP}$ -∅]]]]

The analysis shows that inside the indefinite pronoun, insertion of PluralP forces attraction of DivP, breaking the constituent lexicalized by *-ien*. This lexical item cannot shrink in Spanish because the Elsewhere Principle forces insertion of the desinence *-o*, which contains fewer unmatched features than *-ien* (0 vs., at least, 5). Depending on whether Spanish contains a zero morpheme that lexicalizes AnimacyP, *alguien* is impossible in the plural or, alternatively, it is possible, but needs to be lexicalized by *alguno* for lexical reasons. Determining between these two alternatives requires independent inquiries in the nature of animacy which will be left for further research. In any instance, however, we have shown that *alguien* will never surface in the plural form.

Let us consider now why *alguien* rejects Partitive Constructions. Notice that partitive constructions always require that the coda (the constituent introduced by the preposition *of* in English or *de* in Spanish) is formally plural, something which has been known as the Partitive Constraint (as defined in Partee 1987). Gutiérrez Rodríguez (2008) shows that semantically plural nouns which are singular (such as mass nouns and collectives) are false partitives which, among other properties, allow the prepositional phrase to be *from* in English and *de entre* in Spanish. Thus, the requisite imposed by the presence of the partitive construction is that PluralP is present in the structure. If both the partitive coda and the head of the construction share the same functional sequence (as in (54a)), the incompatibility of *alguien* with partitives follows: the presence of PluralP

breaks the syntactic constituent required by *-ien*, forcing insertion of *un-* and the desinence *-o*. Several arguments show that the right structure for partitives is (54a), and not (54b), where the coda and the quantifier belong to different nominal Fseqs and the quantifier combines with an empty noun phrase, e, coindexed with the partitive complement.

(54) a. [two [of the students]]
b. [two e_i [of the $students_i$]]

First of all, no noun can appear outside the partitive coda in this construction (55a). As Gutiérrez (2008) shows, sequences where two nouns materialize are not partitive codas, but false partitives. Also, the fact that there is only one functional sequence in a partitive construction explains that it is impossible to have another quantifier inside the partitive coda (55b,c): the only quantifier position available in the sequence has already been filled.

(55) a. *two students of the students
b. two of the students
c. *two of many students

Again, Gutiérrez (2008) shows that constructions with two quantifiers (*two of the many students that arrived*) are false partitives; notice that they also allow two nouns (*two students of the many students that arrived*), which make them pattern with the previous group.

There are several conceivable analyses of the partitive constructions under these assumptions, depending on whether movement of the quantifier or selection is assumed, or on whether the preposition is viewed as a case marker or as a relational head with partitive meaning. In any case, once we have established that both the quantity-denoting expression and the partitive coda share the same functional domain, it follows that the quantity-denoting expression is sensitive to the number information contained in the partitive coda.

4.4. Why *quién* allows the plural

At this point, the question is why *-ien* can combine with the plural in interrogative contexts. There are two syntactic properties of wh-words that differentiate them in their formal behaviour from indefinite pronouns and which, to our mind, are universal.

The first one is that wh-words, even in the singular, allow for a plural interpretation. This can be straightforwardly shown by symmetric predicates, which, as noticed before, need to combine with collectivities, groups or masses. The sentences in (56a,b) are ungrammatical because the only argument of the verb is a singular noun, as shown by verbal agreement in each language. In contrast, if the DP is substitituted by a wh-word in the singular, the sentence is grammatical, although this wh-word is still singular, as agreement makes explicit (56c,d).

(56) a. #John meets in this room every day.

b. #Juan se amontona allí.
Juan SE piles up there
c. Who meets in this room every day? (possible answer: John and Mary).
d. ¿Quién se amontona allí? (possible answer: Juan, Pedro y Luis).
who SE piles up there?

Remember that *alguien* cannot combine with collective predicates; in other words, *alguien* behaves formally like a single count entity, but not *quién*, although they both trigger singular agreement.

The wh-word cannot be considered a collective noun, that is, a formally singular noun which denotes a semantically plural entity (like *army* in many languages). Notice that collective nouns in singular can combine with the preposition *entre*, 'among', in Spanish (57a). However, *quién* cannot combine with this preposition, showing that it does not pattern with collective nouns (57b). In this context, the plural form of the wh-word needs to be used to satisfy the plural requisite of the preposition (57c).

(57) a. Dan vivió entre el pueblo pirahã.
Dan lived among the people Pirahã
'Dan lived among the Pirahã people'.
b. *¿Entre quién vivió Dan?
among who lived Dan?
'Among who did Dan live?
c. ¿Entre quiénes vivió Dan?
among who.pl lived Dan?

The problem is how to account for the fact that *quién* is formally a singular, not a collective noun, but can combine with wh-words. Our proposal is that this shows that the number value inside a wh-word is not used to calculate how many atoms of a particular entity there are in the context of discourse. Notice that even if we use *quién* in the plural form, there is no entailment that the question requires a plural answer (58). In a situation where only one person came to the party, the question in (58a) can be answered by (58b) without any infelicity. The use of the plural in the wh-word may show some expectation of the person asking some information, but it is definitely not presupposed in the question itself, as opposed to other pieces of information, such as that the entity she is asking about needs to be a human (cf. (58c), where the answer is infelicitous).

(58) a. ¿Quiénes vinieron a la fiesta?
who-pl came-pl to the party?
b. Mary.
c. #Tres pizzas.
'Three pizzas'.

The fact that number information is contained in the wh-word (as shown by agreement), but that it is not computed as part of the entailment of the sentence and can be ignored by symmetric predicates is captured if inside the wh-word the PluralP does not attract the DivP. In this situation, the information about plural is not able to calculate the number of atoms available in a particular context. In other words, the number projections inside the wh-word are weak.

The second property of wh-words is that they require an exhaustive identification of the entity bound by the interrogative. The question in (58a), uttered in a context where Mary, John and Paul came to the party, requires the answer in (59a); the partial answer in (59b) is infelicitous, as it does not identify exhaustively all the entities in the domain of the discourse. In other words, the answer to a wh-word requires that all the entities that satisfy a particular description (in this case, having come to the party) are identified, without exceptions.

(59) a. #Mary.
b. Mary, John y Paul.

Horvath (2007) argues that the exhaustive identification is made possible by a particular quantifier which, when present, is attracted by an Exhaustive Identification Focus head. Hungarian makes it possible to see when this movement has taken place, and in consequence when the exhaustive identification quantifier is present in the structure, because, when the Exhaustive Focus position is filled, verbal modifiers (which otherwise must appear to the left of the verb) need to occur to the right. The contrast in (60) shows that, if the niece is interpreted as the exhaustive focus of the sentence, the preverbal modifier *be-* needs to be realized post-verbally.

(60) a. Be-mutattam Jánost az unokahúgom-nak
in-showed.1sg Janos.acc the niece.my-dat
'I introduced John to my niece'
b. AZ UNOKAHUGOM-NAK mutattam be Jánost.
the niece.my-dat showed.1sg in Janos.acc
'It was to my niece (and nobody else) that I introduced Janos'
c. * AZ UNOKAHUGOM-NAK be-mutattam Jánost
the niece.my.dat in-showed.1sg Janos.acc
[Horvath 2007: 115]

The sentences in (61) show that the wh-word occupies the exhaustive focus position, and it forces the verbal modifier to be to the right and is not compatible with another exhaustive focus (61c).

(61) a. Ki-nek mutattad be Jánost?
whom-dat showed.2sg in Janos.acc?
'To whom did you introduce Janos?'

b. * Ki-nek be-mutattad Jánost?
c. *Ki-nek Jánost mutattad be?
whom-dat Janos.acc showed.2sg in?
(Intended: 'To whom did you introduce Janos, and to nobody else?')

Indefinites can never be in exhaustive focus; Hungarian shows evidence of this given the syntactic behaviour of *vala-ki*, equivalent to *someone* and *alguien*. (62) shows that it cannot occupy the exhaustive identification focus, as the verbal modifier needs to occupy the preverbal position.

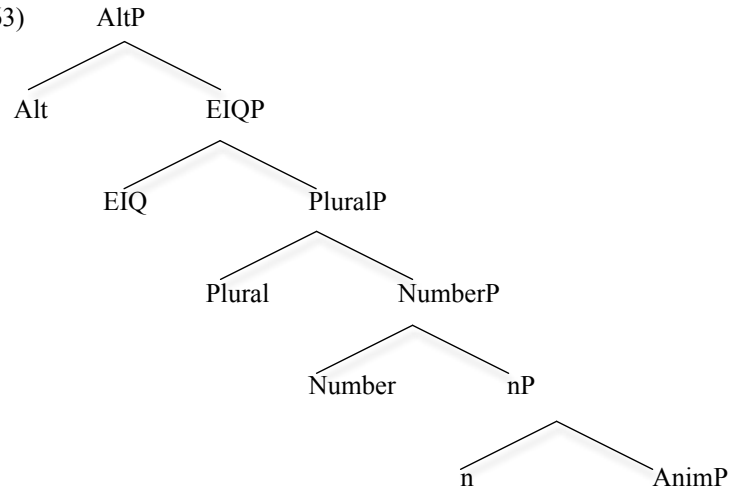
(62) a. Vala-ki-nek be-mutattad Jánost.
some-who-dat in-showed.2sg Janos.acc
'You introduced Janos to someone'.
b. *Vala-ki-nek mutattad be Jánost.
(Intended: You introuced Janos to someone, and to nobody else).

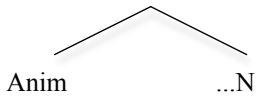
This shows that wh-words contain an exhaustive identification quantifier which allows them to move to the exhaustive focus position, while indefinites do not contain it and are, therefore, banned from occupying that position.

There are two syntactic differences between the indefinite and the interrogative, then. The first one refers to the syntactic projection of number, which is not able to attract the DivP, as shown by the behaviour that the wh-word displays with respect to symmetric predicates. The second one refers to the presence of an exhaustive identification operator in wh-words, but not in the indefinites, as shown by their different behaviour in exhaustive focus. We will show now how these differences explain that *quiénes*, as opposed to *alguien*, allows combination with the plural.

The structure of the wh-word is the one presented in (63), where the projection only expands to AltP.

(63)

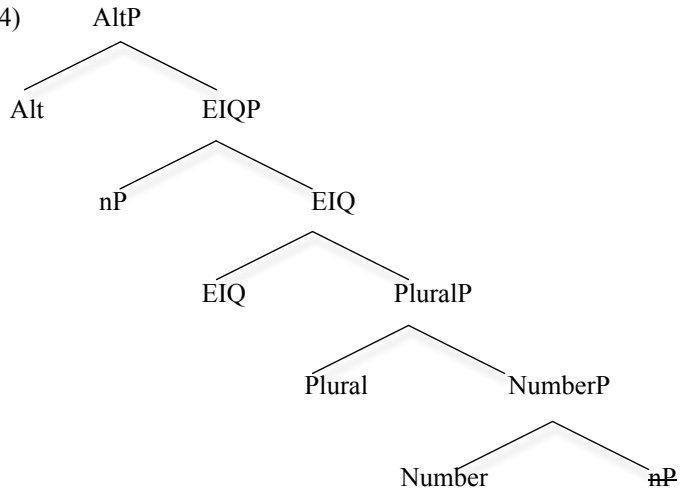




As can be seen in (63), we propose that the right position for the Exhaustive Identification Quantifier (EIQ) is below AltP and over the number projections. The evidence for this comes from the fact that number information does not need to be exhaustive, while exhaustivity needs to take into account the number information. Remember that in a context where more than one entity satisfies a description, exhaustive focus requires the entity to be plural. In contrast, in generic contexts, exceptions are allowed, in such as way that *The lions in Africa are brown* is not considered to be false if some lions in Africa do not fit that description.

In order to satisfy its semantic role, the EIQ needs to have access to the index of identity of the entity, which is contained in nP. We propose that this is performed via a syntactic movement. The projection nP, with everything that it contains, is attracted by EIQP and merges in its specifier, as presented in (64).

(64)



Even if plural is present in the structure, it will not attract DivP, because, as shown before, the number value of the structure is not calculated by computing the atoms provided by DivP.

After the movement in (64) takes place, the constituent that *-ien* is associated to in the lexicon is destroyed, but notice that the Superset Principle allows *-ien* to shrink and get inserted in nP, which contains the features n, Animacy, Div and N. In this context, *-ien* is not blocked by the desinence, because the desinence corresponds to a smaller unit, and therefore cannot be inserted in nP. Thus, *-ien* lexicalizes nP and all that it contains. As the nP does not form a constituent with AltP, *-ien* cannot lexicalize this head, and the lexical item *qu-* is inserted to perform that task. Finally, in case PluralP is present, the lexical item *-s* is

inserted. The resulting morphological structure is represented in (65). We assume that, as in Hungarian, the EIQ is realized by a zero morph.

(65) $[_{AltP}$ qu- $[_{EIQP}\ [_{nP+Anim+Div+N}$ -ien] ∅ $[_{PlP}$ -es]]]]

In consequence, *-ien* can combine with the plural form in the interrogative because in this context the PluralP does not attract DivP, and, as opposed to the indefinite, there is an Exhaustive Identification Quantifier which attracts nP, allowing *-ien* to shrink and lexicalize this constituent.

4.5. Extending the analysis to Hungarian.

We will show at this point that the proposed structure and the operations mentioned can explain other patterns attested in different languages. Here we will concentrate on Hungarian.

The difference between Spanish and Hungarian is that in this second language both the indefinite and the interrogative, morphologically related to each other, allow the plural morpheme *-k*.

(66) a. vala-ki-k
some-who-pl
'some people'
b. ki-k
who-pl
'which people'

This difference in the behaviour of the indefinite pronouns can be straightforwardly related to a general difference between the lexicon of Spanish and that of Hungarian. This second language, as opposed to Spanish, does not have the equivalent of desinences in its lexicon. As it is known, Hungarian does not have lexical oppositions based on gender in pronouns, and does not differentiate the gender information of animates by a specific morpheme associated to the noun class. (67a) is the pronoun for both he and she, while (67b) can refer to either a male or a female entity. When Hungarian differentiates between a masculine and a feminine lexically, this is always performed by using two different roots, as in (67c), or by adding the free form *nö*, which is otherwise used as a politeness pronoun (67d). Finally, Hungarian adjectives do not contain morphemes that agree with the nouns in gender, as shown in (67e).

(67) a. ö
b.orvos
doctor (male or female)
c. fiú ~ lány
boy ~ girl
d. barát ~ barat+nö
boyfriend ~ girlfriend

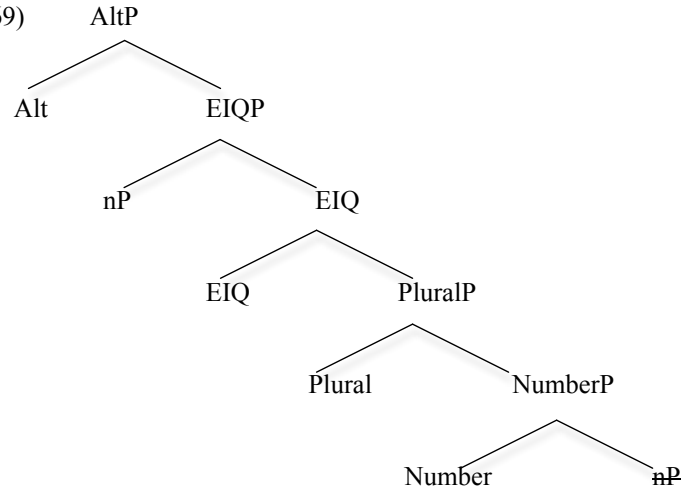
e. a magas orvos
the tall doctor (female or male)

The comparison with Spanish shows that Spanish can perform all these operations by substituting the desinence attached to the noun and the adjective. This shows that Hungarian does not contain a lexical item which spells out the area that the desinence spells out in Spanish. Although Hungarian differentiates count and mass nouns and has meaning differences related to gener, these differences are not marked by specific lexical items.

(68) a. él - ella
b. doctor ~ doctor-a
male doctor ~ female doctor
c. chico ~ chica
boy – girl
d. el doctor alto ~ la doctora alta
the tall (male) doctor ~ the tall (female) doctor

Let us go back to the structure for the indefinite, containing the PluralP. The syntactic structure, both in Spanish and in Hungarian, is the one in (69).

(69)



In Spanish this structure cannot be lexicalized by *-ien* because there is a smaller lexical item to materialize DivP, the desinence. However, Hungarian lacks this lexical item, so the Elsewhere Condition would not block use of the same lexical item. In other words, *ki*, partially equivalent to *-ien*, can shrink to the DivP. The higher projections are lexicalized in Hungarian with *vala-*. The other details are like in Spanish.

(70) $[_{PPI+\exists P+AltP}$ vala $[_{PluralP}$ $[_{DivP}$ ki] -k $[_{NumberP+nP}$ ∅]]]]

Thus, absence in Hungarian of a smaller lexical item that competes with *ki* allows this language to have plural forms in its human indefinite pronoun.

4.6. Something about English: unshrinkability

We would like to finish with a note about English and, in general, the languages where the interrogatives do not allow a plural form in any case. Although we will not provide here a full fledged analysis of all these languages, we will make some remarks about English that hopefully will point to a fruitful future analysis of this third pattern.

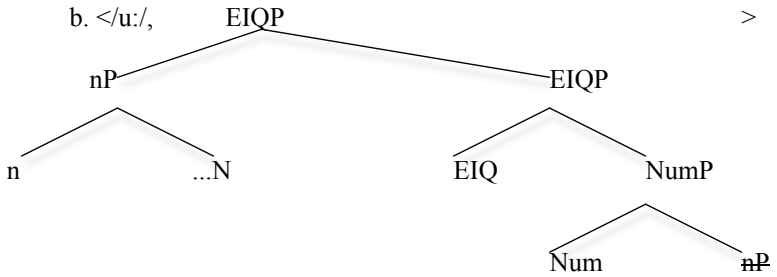
When comparing Spanish, Hungarian and English, the first thing that attracts attention is the fact that, while in the previous two languages the indefinite and the interrogative share morphemes, in the last language none of the morphemes is shared.

(71) a. alg-**ien**, qu-**ien**
b. vala-**ki**, **ki**
c. somebody / someone, who

This strongly suggests that in English, the lexical entry for the morphemes involved in the interrogative is associated to a tree in which the movement operation has already been performed. In other words, the fact that the indefinite and the interrogative share some syntactic features (as they are both animates, both count and both cannot combine with overt nouns) but do not share any lexical item suggests that the lexical items are different because they also lexicalize the head to which that portion of the tree has moved. In the case of the interrogative, then, assuming than *wh-* corresponds to the AltP, *-o* corresponds to the constituent formed by the EIQP with nP in its specifier.

(72) a. </h/, Alt>

b. </u:/,



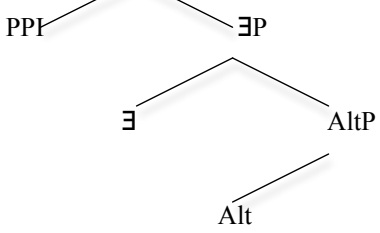
As can be seen in the entry for /u:/ in English, the tree to which the lexical item is associated contains a specifier, plus a head and its complement. Inside the constituent to which the lexical item is associated, the base copy of the element

that has moved to the specifier position is located. Our proposal is that this kind of entry cannot be shrunk, which guarantees that PluralP cannot be found in the syntactic structure.

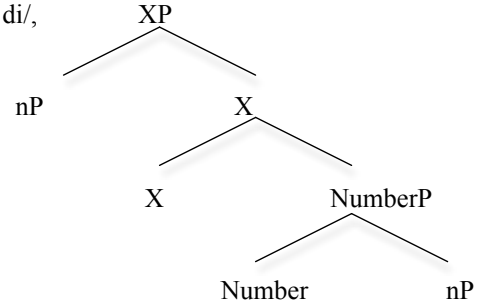
The reason for this situation is the following. It is generally assumed that, when multiple copies are present in the structure, only one of them can be lexicalized, generally the hierarchically higher one. At the same time, given the Anchor Condition, lexical items can only shrink if they are anchored to the lowest head of the constituent that they lexicalize. If the structure in (72) was lexicalized by two or more lexical items, the copy of the nP contained in the domain of EIQ would make lexical insertion crash. Given the general ban on materializing copies multiple times, this nP could not be lexicalized twice. However, if this nP is marked as unable to be associated to lexical material, no lexical item can materialize the domain of EIQ, as no lexical item can anchor to its lower node(s), as shrinking requires. The only option in this situation is to lexicalize the whole constituent with a single lexical item which, by the reasoning just presented, is not able to shrink. This makes it impossible to introduce intermediate heads in the structure, as this would force the lexical item to shrink.

The same situation would arise in the case of the indefinite if we assume that in English, as opposed as in Spanish and Hungarian, the nP is attracted by a higher head in the indefinite pronoun.

(73) a. </s□m/, PPIP >



b. </b□di/, XP >



The generalization would be, in general, that a lexical entry that involves a moved constituent cannot be shrunk in any situation.

As can be easily understood, it would be necessary to make explicit the nature of XP in this structure, determine the nature of the movement operation involved and explore the other differences between English and the other languages studied which follow from this. All these topics will be left for further research.

5. Advantages of Phrasal Spell Out for morphological theory

Now that we have shown that PSO is allowed by the present theory and, more importantly, that there is empirical evidence for its existence, we would like to show how it can be used to solve a long-standing controversy in linguistics. In Item-and-Arrangement morphology (see, for example, Hockett 1947) the systematic aspects of form-meaning association inside words are captured by proposing that lexical items are morphemes. As noticed very early in the proposal, however, this theory runs into trouble when the grammatical differences that are normally expressed by adding morphemes involve suppletive or allomorphic changes in the stem of the word. Consider, in this line, (74a), as opposed to (74b).

(74) a. Present: *move*; Past: *moved*
b. Present: *go*; Past: *went*

As Nida (1948) points out, while the first case could be analyzed as the addition of a past morpheme *-ed* to the verb, trying to apply this method to the second case produces problems. The parallelism with (74a) would force (74b) to look as represented in (75b), as Hockett did.

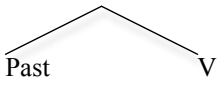
(75) a. [move] + [ed]
b. [went] + [∅]

The problem of the segmentation in (75b) is that the contrast between the present and the past of *go* is performed by a zero morpheme, which, by definition, does not produce form oppositions inside the paradigm, while the object that shows the form opposition (*went*) is considered identical to *go*, with form changes being forced by the presence of the zero morpheme. On the other hand, an Item-and-Process theory or a Word-and-Paradigm one could explain (74b) by proposing that a rule has transformed the verb stem or the word without proposing any internal segments inside the word.

Item-and-Process and Word-and-Paradigm theories have the problem, however, that in sequences of clearly segmentable affixes, the relationship between these affixes is not made explicit inside the word. In a word like *readability*, for example, the hierarchical relationship between *–abil-* and *–ity* cannot be captured, leaving unexplained the fact that the word is understood as the property of being readable and not as the possibility of having the property of reading. Thus, the two sets of theories encounter different problems and are able to explain only part of the data.

PSO solves this tension, giving us the best of the two possible worlds. Notice that the two words in (74) could be analysed by basically the same structure; the difference arrives when lexical insertion takes place, as a particular lexical item is able to lexicalize a phrase headed by Past in the verb *go*, but not in the verb *move* (76). The lexical item *went* would be inserted in PastP (the constituent containing both heads), while *-ed* and *move* are inserted in the terminal nodes Past and V.

(76) a. PastP



b. move <---> {V}
-ed <---> {Past}
went <---> {Past, V}

From the point of view of the tree, the system is necessarily 'Item-and-Arrangement' and differentiates between discrete units that compositionally construct the semantic and formal aspects of the structure. However, the lexicalization procedure can be 'Item-and-Process', because whole phrases may be lexicalized by a single unit (as is the case with *-ien* or with *went*), not materializing the different ingredients of the word as separate segments. The syntactic heads, once we allow for PSO, can be submorphemic entitities, that is, entities which are included inside a morpheme but not necessarily are identical to it. As Nida (1948) himself puts it in his discussion of suppletivism, "practically all the features of concord, government and cross-reference could be treated on a submorphemic level". PSO provides the formal mechanism to implement this intuition, and helps to bridge the gap between theories that assume the existence of morphemes and those that don't.

References

Abels, Klaus and Peter Muriungi. 2008. The focus particle in Kîîtharaka: Syntax and Semantics. *Lingua* 118: 687-731.

Alexiadou, Artemis and Kirsten Gengel. to appear. Classifiers as Morphosyntactic Licensors of NP ellipsis: English vs. Romance. *Proceedings of NELS 39*.

Baker, Mark. 2003. *Lexical categories*. Cambridge University Press, Cambridge.

Bernstein, Judy. 1993. *Topics in the Syntax of Nominal Structure across Romance*. Ph.D. thesis, CUNY.

Blake, Barry J. 2000. *Case*. Cambridge University Press, Cambridge. Second edition.

Borer, Hagit. 2005. *In Name Only*, vol.1 from *The Exoskeletal Trilogy*, Oxford University Press, Oxford.

Caha, P. (2009): *The Nanosyntax of Case*. Ph.D. thesis, CASTL, Universitetet i Tromsø.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and Beyond. The Cartography of Syntactic Structures*, vol. 3., edited by Adriana Belleti. Oxford University Press, Oxford, pp. 104-131.

Chomsky, Noam. 2008. The biolinguistic program: Where does it stand today? Ms., MIT.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry* 36: 315-332.

Citko, Barbara. 2005. On the nature of merge: external merge, internal merge and parallel merge. *Linguistic Inquiry* 36: 475-496.

Corbett, Greville. 1991. *Gender*. Cambridge University Press, Cambridge.

Elbourne, Paul. 2002. *Situations and Individuals*. Ph. D. dissertation, MIT.

Fábregas, Antonio. 2007. An exhaustive lexicalization account of directional complements. In *Tromsø Working Papers on Language and Linguistics: Nordlyd 34.2, Special issue on Space, Motion, and Result*, edited by Monika Basic, Marina Pantcheva, Minjeong Son & Peter Svenonius (eds.), University of Tromsø, Tromsø, pp. 165–199.

Fábregas, Antonio & Isabel Pérez. 2008. Gender agreement on adverbs in Spanish. *Journal of Portuguese Linguistics* 7: 25-47.

Fox, Danny. 2000. *Economy and Semantic Interpretation*. MIT Press, Cambridge.

Gärtner, Hans M. 2002. *Generalized Transformations and Beyond: Reflections on Minimalist Syntax*. Akademie Verlag, Berlin.

Grimshaw, Jane. 2005. *Words and Structure*. CSLI, Stanford.

Gutiérrez Rexach, Javier. 2003. *La semántica de los indefinidos*. Visor, Madrid.

Gutiérrez Rodríguez, Edita. 2008. *Rasgos gramaticales de los cuantificadores débiles*. Ph.D. dissertation, Universidad Complutense de Madrid.

Harley, Heidi and Rolf Noyer. 1998. Licensing in the non-lexicalist lexicon: nominalizations, Vocabulary Items and the Encyclopedia. In *MIT Working Papers in Linguistics 32: Papers from the UPenn/MIT Roundtable on Argument Structure and Aspect*; edited by Heidi Harley. MIT, Cambridge, pp. 119–137.

Heim, Irene. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts.

Hockett, Charles. 1947. Problems of morphemic analysis. *Language* 23: 321-343.

Horvath, Julia. 2007. Separating focus movement from focus. In *Phrasal and Clausal Architecture*, edited by Simin Karimi *et al*. John Benjamins, Amsterdam,108-145.

Kamp, Hans. 1981. A theory of truth and semantic representation. In *Formal Methods in the study of language*, edited by J. Groenendijk *et al*. Amsterdam Center, Amsterdam, 277-322.

Kratzer, Angelika and Junko Shimoyama. 2002. Indeterminate pronouns: the view from Japanese. Paper presented at the *3rd Tokyo Conference in Psycholinguistics*, Tokyo.

Laca, Brenda and Liliane Tasmowski. 1996. Indefini et quantification. *Recherches linguistiques de Vincennes* 25: 107-128.

Martí, Luisa. 2008. The semantics of plural indefinite noun phrases in Spanish and Portuguese. *Natural Language Semantics* 16: 1-37.

McCawley, James. 1968. Lexical insertion in a transformational grammar without deep structure. *Papers from the Fourth Regional Meeting Chicago Linguistic Society*. University of Chicago, Chicago, 71-81.

Neeleman, Ad and Kriszta Szendröi. 2007. Radical pro-drop and the morphology of pronouns. *Linguistic Inquiry* 38: 671-714.

Nida, Eugene. 1948: The identification of morphemes. *Language* 24: 414-441.

Partee, Barbara H. 1987. Noun phrase interpretation and type shifting principles. In *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. edited by J. Groenendijk *et al.*, Foris, Dordrecht, 115-143.

Picallo, Carme. 2006. On gender and number. Ms., Autonomous University of Barcelona.

Ramchand, Gillian. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge University Press, Cambridge.

Starke, Michal. 2001. *Move Dissolves into Merge: A Theory of Locality*. Ph.D. thesis, University of Geneva.
Svenonius, Peter. 2005. Extending the extension condition to discontinuous idioms, *Linguistic Variation Yearbook* 5: 227-263.
Weerman, Fred and Jacqueline Evers-Vermeul. 2002. Pronouns and Case. *Lingua* 112: 301-338.