QUANTIFYING COMPOUNDS

Juan Carlos Odriozola Pereira
UPV/EHU

0. Introduction[1]

This paper is concerned with N-N sequences where the second nominal item, which is usually understood as the head of the construction, actually instantiates a quantification of the first noun. The word 'compound' in the title of this paper is the label used in the bibliography for such sequences in certain languages. Here we focus on Basque N-N sequences in phrases like *bi ogi zati* lit. 'two bread piece' (two pieces of bread), *ogi pitin bat* lit. 'bread bit one/a' (a bit of bread), *bi mutil multzo* lit. 'two boy set' (two sets/groups of boys) and *mutil pilo bat* lit.'boy lot one/a' (a lot of boys).[2] Depending on the distribution of quantifiers and (other) determiners, these sequences may take a quantifying reading. Hereafter we will be speaking about quantifying N-N sequences, and quantifying and quantified nouns, without regard to whether in each particular case the quantifying reading appears.

Section 1 analyses these sequences from two different points of view: that of Xº level compounds, and that of the general structure of Basque noun phrases. Section 2 explains the distribution of Basque quantifying and quantified nouns in N-N sequences, N-P-N sequences and measure phrases as a starting point for applied linguistics activities such as language processing and languages for special purposes.

1. Are quantifying N-N sequences compounds?

A revision of the bibliography on what constitutes a compound would exceed the scope of this paper.[3] Instead, we will review certain criteria concerned with the

[1] This research is supported by grants no. HUM2004-05658-C02-01 and EHU06/16, from which a few examples from Economics and Natural Sciences are taken, respectively. Some general data of Georgian come from the "Basque-Georgian Comparative Morphology" project directed by professor Tamar Makharoblidze of the Tbilisi State University of Georgia. I would like to thank José María Arriola, Xabier Artiagoitia, Xabier Bilbao, Luis Eguren, Edurne Etxebarria, Deanie Johnson and Ainara Ondarra for their valuable comments and suggestions. All errors are, as always, my own.

[2] In footnotes, data of a pre-IndoEuropean language like Georgian will be compared to data from a pre-IndoEuropean language like Basque. Some data from Germanic languages (English and German) and Romance languages (Catalan and Spanish) will also be provided. Our purpose is to sketch a starting point for research lines within typology, linguistic theory and applied linguistics.

[3] The bibliography on several languages (Becker 1993, Buenafuente 2007, Marchand 1960: §II, Lieber 1993, Pérez-Gaztelu et al. 2004, Rainer & Varela 1993) provides some well-known criteria such

Basque quantifying N-N sequences that appear in (1-2) in determiner/quantifier phrases:

(1) a bi ogi zati
two bread piece
'two pieces of bread'

b ogi pitin bat
bread bit one/a
'a bit of bread'

(2) a bi mutil multzo
two boy set
'two sets of boys'

b mutil pilo bat
boy lot one/a
'a lot of boys'

It is clear that in (1-2) the second noun instantiates a kind of quantification of the first. Basque is a head-final language provided with postpositions, but neither a postposition nor any other element is needed between the quantified and the quantifying noun, which contrasts with the case of English and the Romance languages. (1-2) pattern apparently with other Basque N-N sequences included in noun phrases of the type of *bi ogi-denda* lit. 'two bread shop' (two bakeries) and *mutil-dantza bat* lit. 'boy dance one/a' (a dance only performed by boys). (1-2) and *ogi-denda* and *mutil-dantza* all have been explored together and taken as dependency compounds in the bibliography, the noun to the right being taken as the syntactic and semantic head of the construction. However, unlike *ogi-denda* or *mutil-dantza,* the sequences in (1-2) almost never appear in dictionaries as lexical entries. Furthermore, the quantifying reading of (1-2) contrasts with non-quantifying readings in sequences like '*bi ogi zatiak*' lit. two piece bread-the' (the two pieces of bread), *bi mutil multzoak* 'two boy set-the' (the two sets/groups of boys). Finally, the noun *pitin* 'bit' is not allowed in article-bearing phrases of the type **ogi pitina* lit. 'bread bit-the'. All these data need an analysis that is more sophisticated than the one where a Xº N-N compound has a left element that depends semantically and syntactically on the lexical element situated to the right.

Section 1.1 explores the thematic/semantic relationships between the two elements, assuming that they are fully nominal. In section 1.2, we leave aside the nominal compound approach and consider instead the functional quantifier characteristics of the nouns to the right in these sequences.

1.1. Relationships between the elements of quantifying N-N sequences

1.1.1. Thematic relationships

The bibliography on English tends to distinguish two main types of compounds: the so-called root compounds like *skyline* and the so-called synthetic compounds like *skyscraper*. Two crucial characteristics are described for synthetic compounds. Marchand (1960: §1.2.3) and Bauer (2001: §4.3.2, §5.2.8, §6.5) emphasize that the noun *scraper* is not exactly derived from the verb *to scrape* and seems to be cre-

---

as the Xº level of the whole structure, non-compositional meaning, unspecific reading of the non-head constituent and a stress structure that is different from that of the phrases. Regarding N-N sequences, not all of them are taken as compounds in all the works concerning English, whereas until now all Basque N-N sequences have been taken as compounds. Regarding compounds in general, the fact is that the set of general criteria mentioned above varies notably inter- and intralinguistically.

ated ad hoc as a pseudo-basis for *sky*. Basque has a word formation case that is still clearer (Euskaltzaindia 1987: 44-45, Euskaltzaindia 1991: §25-60): the suffixes *-gin* (cif. *egin* 'to make'), *-gile* (cif. *egin* 'to make', *egile* 'maker') and *-dun* (cif. *duen* 'which has') give rise to *zapatagin* 'shoemaker' (cif. *zapata* 'shoe'), *langile* 'worker' (cif. *lan* 'work'), *bizardun* 'bearded' (cif. *bizar* 'beard'), although *-gin*, *-gile* and *-dun* only function as suffixes in today's Basque. Therefore, the label of 'synthetic' is initially related to sequences that are on the borderline between the compounding and deriving processes,[4] where the head of the sequence is close to being an affix.

Professor Tamar Makharoblidze (p.c.) reminds us that the term 'synthetic' comes from typology. In the case in hand, typology can be furthered through morphological productivity (Bauer 2001: §1.1, 2.5.), or the exploration of the ability for forming new words. That is, languages can be taken as (more) analytic when they tend to express new notions by means of syntax, and more synthetic when they tend to use word formation. Therefore, *skyscraper* is (taken as) more synthetic than *skyline* because the former is less syntax-like than the latter. Following Marchand, *skyline* (and not *skycraper*) seems to be the primary compound type that arises from combining two fully independent common nouns.

The other crucial point is that the noun to the left is taken as an argument of the head of synthetic compounds (Bauer 2001, Grimshaw 1990: §3.4, Lieber 1993, Marchand 1960). Let us informally remember that the argument structure framework is based on the idea that some lexical-conceptual features of a given item are necessarily understood at all linguistic levels and tend to need be expressed either in morphology or in syntax (Grimshaw 1990, Gràcia 1994, Gràcia et al. 2001). Although the bibliography does not speak about the lexical-conceptual requirements of quantifying nouns (Gràcia et al 2000: §1.4.2.6, Pustejovsky 1998:§ 8), argument structure theory is closely related to thematic theory. Outside the strict argument structure framework, thematic relationships between nouns are explored in works that pay attention to what in general has been taken as possessive thematic relationships. In short, Castillo (2001) and Odriozola (2006b) show that in noun pairs such as *ogi/zati* and *mutil/multzo*, the fulfilment of the requirements of *zati/multzo* by *ogi/mutil* is almost as obvious that in the case of *sky* and *scrape(r).*

So from a lexical point of view, what we call a 'quantifying noun' seems to impose a thematic relationship on what we call a 'quantified noun'.

1.1.2. The items in Basque quantifying N-N sequences have a meronymic relationship

The main semantic/syntactic criteria that the bibliography uses to classify Basque N-N sequences has been outlined by Marchand (1960: §2.1.1, §2.1.2, §2.13), who describes a general pattern for English N-N compounds such as *skyline*, where the

---

[4] The term 'synthetic' is related to the particular problem of the borderline between composition and derivation in the standard processes of several languages. Grácia et al. (2001) and Pérez Gaztelu et al. (2004) point out that the Spanish process in *des-* + *grano* 'grain' + -ar > *desgranar* 'to thresh' is a parasynthetic process in that **granar* is not available in the Spanish lexicon. Basque counterparts featuring free derived agents and event nouns are also seen as synthetic (Azkarate 1990: 267-285, Pérez et al. 2004).

right constituent *line* is what he calls the *determinatum*, since the compound expresses a kind of *line*, and the left constituent, *sky*, which is the determinandum. The general criteria used in the Basque bibliography is the (lack of a) dependency relationship between the two constituents of the compound. Azkarate (1990), Odriozola & Pérez-Gaztelu (2002), and Pérez-Gaztelu et al. (2004) basically distinguish Basque N_1-N_2 sequences where N_1 syntactically and semantically depends on N_2[5] from sequences where a dependency between N_1 and N_2 cannot be confirmed.

Semantically, a dependency compound like *itsasgizon* 'seaman' denotes a subclass within the class denoted by the N_2 *gizon* 'man'. Syntactically, the head imposes its categorical features on the compound.[6] Under this determinatum/determinandum pattern, both root compounds like *skyline* and synthetic compounds like *skyscraper* are dependency compounds. By contrast, both *ogi zati* lit. 'bread piece' and *mutil multzo* lit. 'boy set' clearly differ from Basque root compounds (*itsasgizon* 'seaman') and Basque synthetic compounds (*kale-garbitzaile* lit. 'street cleaner').

The only specification provided for Basque quantifying N-N sequences (Euskaltzaindia 1987: §11.2) is that they are dependency compounds that have a 'special' head or determinatum. One of the main aims of this paper is to show that the concepts of hyponymy and syntactical dependency between nominal elements are not accurate for Basque quantifying N-N sequences. In Catalan, Solé (2002) was interested in determining —simplifying her large set of data and using her terms in a very broad sense— whether a given element expresses a subclass of a collective noun (in a hyponymic relationship) or whether it expresses a part of the collective noun (in a meronymic relationship). Basque quantifying N-N sequences follow two general patterns. In the first, nominal elements like *zati* 'piece' denote a part of (mass collective) nouns like *ogi* 'bread'. In the second, items like *multzo* 'set' express a (collective) whole and quantified nouns like *mutil* 'boy' denote a (countable) part. In both, the relationship between the nouns is not hyponymic. Making a simplification, we assume that all the relationships involved in Basque quantifying N-N sequences are meronymic, unlike what has been described in the literature.

1.2. To what extent quantifying N-N sequences are not compounds

1.2.1. Quantifying nouns and quantifiers

Even if instead of a hyponymic relationship related to compounds we assume that quantifying N-N sequences have a meronymic relationship, the fact of quantification

---

[5] The head parameter and the complement/head relationship (that can be seen in both synthetic compounds and quantifying N-N sequences) is rather misleading: a head-initial language such as English patterns with a head-final language like Basque in that both tend to produce right-headed compounds, whereas a head-initial language like Spanish tends to produce left-headed compounds. This is so for each language, whatever the relationship between compound elements. Of course, the two optional positions for modifiers could recover the X'-grammar inside the X^o-level.

[6] In sequences with two Ns, no overt imposition of the category of the head can be confirmed. However, Odriozola & Pérez-Gaztelu (2002) note that the dependency can also be checked in the subcategorizations of both nouns: it is obvious that the subcategory of the compound is matched with the subcategory of the head. That is, given that *itsaso* is not human and *gizon* is human, the compound takes the subcategorization features of the constituent that is taken as the head —i.e. from *gizon*.

itself is not very well captured by assuming a pattern of two nominal elements that are considered to be fully lexical.

This section leaves aside that point of view and aims at relating the optional quantifying reading to a different structure having a functional projection headed by a quantifier phrase. This is not a theoretical paper. Instead, our aim is to account for all descriptive data by making some minimal assumptions about both the structural relationships and the functional/lexical nature of the different elements. Briefly, we assume that phrases are headed by a lexical or functional element that takes a phrase as complement. The concept of extended projection captures the fact that every lexical projection is a complement of a functional category. Both the functional head and the lexical head give some or all of their features to the whole (functional) projection. The position of the complement parametrically varies from a head-initial language like English to a head-final language like Basque. Besides the complement, it is a matter of fact that projections may have a specifier place located to the left in all languages and that it is usually occupied by a phrase.

Borer (2005: §6.1., §6.2) points out that the English article *the* can appear both with numerals like *one* or *three* and with weak quantifiers like *few,* whereas the indefinite article *a/an* cannot. Therefore she claims that *a* and *the* compete for being a head that must be projected by taking a complement that can be either a noun phrase or a quantifier phrase. Numerals and weak quantifiers would be heads of their own projection or head/phrases in the specifier of a noun phrase.

Basque measure phrases,[7] numerals and some weak quantifiers appear to the left of the noun, whereas some other weak quantifiers and strong quantifiers appear to the right. Artiagoitia (2002) points out that left quantifiers are phrase-like or clearly phrases whereas right quantifiers seem to be heads. So English numerals and weak quantifiers and Basque left quantifiers could be quantifier (phrases) located in a left specifier of a noun phrase. English *the* and *a* and Basque right quantifiers would be heads of a quantifier/determiner phrase that would take either another determiner/quantifier phrase or a noun phrase as complement. English is a head-initial language and has all determiner/quantifiers to the left of the noun phrase, i.e., quantifier phrases or quantifiers in the (left) specifier of the noun phrase and quantifier heads that take a right noun phrase as a complement. Basque is a head-final language with quantifier phrases or quantifiers in the (left) specifier of the noun phrase and quantifier heads that take a left noun phrase as complement. Finally, the Basque attached article *–a* 'the', and the item *bat* 'one/a' appear to the right of the noun, as a functional head of their own projection. Both the non-quantified and the quantified reading of quantifying nouns will be viewed in this framework, under two different analyses.

Regarding the type of readings, when no overt plural mark appears, Basque *-a* is not a conventional article in that it entails a reading that can be either specific or non-specific for what are commonly accepted as mass nouns: e.g. *ogia* lit. 'bread-the' ((the) bread). The article *-a* is related to specific readings[8] of count nouns: *mutila*

[7] For the different readings of these Basque structures, see Etxeberria (2005), Odriozola (2006b) and Trask (2003).

[8] Nouns are not basically mass or count. See this view in Borer (2005), Castillo (2001), Odriozola (2006b).

lit. 'boy-the'. With both mass and count nouns, *bat* 'one/a' has a non-specific reading. Nouns like *zati* and *multzo* do not change the specificity of the reading when the article appears: *ogi zatia* 'the piece of bread', *mutil multzoa* 'the set' of boys'. Luis Eguren (p.c) has noted that in such cases the readings actually are not quantifying. When *bat* appears, the common non-specific reading of both *zati bat* and *multzo bat* changes to a non-specific quantifying one: *bi ogi zati* 'two pieces of bread', *bi mutil multzo* 'two sets of boys'. Therefore, *zati* and *multzo* quantify a noun phrase when they appear with the determiner/quantifier *bat* 'one/a', but they behave as non-quantifying heads of N-N sequences when the article *-a* is attached. This second option seems to be closer to a (compound) analysis where the quantifying noun is the lexical head.

Zati and *multzo* allow some right strong quantifiers like *bakoitz* 'every' and *guzti* 'all', which have the attached determiner and take a generic reading. Weak quantifiers without an article and numerals are allowed too: *ogi zati gutxi* lit. 'bread piece few', *bost ogi zati* lit. 'five bread piece'. Finally, numerals are allowed in phrases headed by the article *-a*, which as usual avoids the quantifying reading: *bi ogi zatiak* 'the two pieces of bread'.

Crucially, some other quantifying nouns such as *pitin* 'bit' only allow the right weak determiner/quantifier *bat* 'one/a' (3a). The article *-a* (3c), numerals (3e) and weak quantifiers are avoided. Weak quantifiers cannot take the article (3d), nor numerals (3f). However, *pitin bat* is the form that always appears in N-N sequences (3a), whereas *bat* is not allowed with standard quantifiers (3b).

(3) a ogi pitin bat
bread bit one/a
b *ogi gutxi bat
bread little one/a
c *ogi pitina
bread bit-DET
d *ogi gutxia
bread little-DET
e *bost ogi pitin
five bread bit
f *bost ogi gutxi
five bread little
g *ogi pitin
bread bit
'a bit of bread'
h ogi gutxi
bread little
'a little bread'

Therefore, *pitin* only entails a quantifying non-specific reading. Nouns like *pilo* 'lot' also have a unique non-specific quantifying reading, although they may appear either with the article *-a* or with the determiner/quantifier *bat*.

Non-quantifying N-N sequences allow left phrases headed by the attached postpositon *-ren*: *ogiaren zaporea* 'the taste of bread', *mutilen garaiera* 'the stature of boys'. *Zati* and *mulzo* allow these postpositional phrases, whereas *pitin* and *pilo* do not:

(4) a *ogiaren pitin bat
bread-of bit one/a
'a bit of bread'
b *mutilen pilo bat
boy-of lot one/a
'a lot of boys'

Pitin and *pilo* have a quantified non-specific reading in *ogi pitin pat* 'a bit of bread', *mutil pilo bat* 'a lot of boys'. The ungrammaticality of (4a-b) seems to be further evidence for a (complex) determiner/quantifier like *pitin bat, mordo bat* that takes a left noun phrase. However it should be remarked that, unlike *pitin bat*, the

collective noun *pilo* allows both the determiner/quantifier *bat* (8d) and the article *–a* without losing the quantifying reading: *Mutil piloa etorri da* lit. 'boy lot-the has come' 'A lot of boys came' takes the same quantifying reading as *mutil pilo bat etorri da* 'boy lot one/a has come' 'A lot of boys came'.[9]

Finally, regarding the analysis of both Borer (2005) and Artiagoitia (2002), additional evidence of the phrasal nature of the left element can be provided. Although both quantifying and non-quantifying N-N sequences such as *ardi-esne* lit. 'sheep milk' are very common, sequences with more than two elements are almost never allowed in Basque. The constraint concerns both N-Adj (5a) and N-N sequences modifying the head (5b).

(5) a *ardi beltz esnea
sheep black milk-DET
'black-sheep milk'

b *mendi ardi esnea
mountain sheep milk

However, Odriozola (2007) points out that the sequences in (6) are grammatical:

(6) a gazta berde zati bat
cheese green piece one/a
'a piece of green cheese'

b ardi beltz talde bat
sheep black group one/a
'a group of black sheep'

c ardi gazta zati bat
sheep cheese piece one/a
'a piece of sheep cheese'

d mendi txolarre talde bat
mountain sparrow group one/a
'a group of mountain sparrows'

At least the sequences N-Adj in (6a-b) are overtly noun phrases. Crucially, Basque allows these kinds of left components when the right nominal is a quantifying noun.[10] We assume that the ability to allow an overt noun phrase to the left is evidence for the analysis that a Basque quantifier can take a noun phrase as a complement to its left.

So far, what we have called quantifying nouns actually take an optional or an obligatory non-specific quantifying reading in Basque N-N sequences,[11] which in most cases is related to the determiner/quantifier *bat* 'one/a' that follows them. The quantifying reading can be captured by assuming a phrase headed by a (complex) quantifier consisting of items like *zati* 'piece', *pitin* 'bit', *multzo* 'set' and *pilo* 'lot' plus an additional standard determiner/quantifier. The non-quantifying reading is related

[9] Generally speaking, the article -a is available only for the non-quantifying reading of weak quantifiers and yet *pilo* takes a weak quantifying reading with the article *-a*, a phenomenon for which we don't have a clear explanation.

[10] Some other classes of nouns allow N-N elements to the left, but they all indicate a specific relationship with the left element. Estopà (1996) describes two kinds of complex determiners: those that instantiate a quantification of the noun and those headed by nouns such as *typus* 'type' that express a graduation of the nouns. In any case, it should be stressed that English and Romance nominal quantifying sequences follow the pattern numeral + determiner noun + preposition + noun, not the N-N sequences.

[11] As said before, Romance languages do not usually have quantifying N-N sequences but have N-P-N sequences of the type (in Catalan) *una mica de pa* and *un grup de noies,* which are similar to the English counterparts *a piece of bread* and *a lot of boys*. Preceding the line of this paper, Estopà (1996) sees in this a kind of complex determiner.

to an Xº-level N-N compound that is headed by what we have called a 'quantifying noun' and that in this case should be assumed to be fully lexical.

1.2.2. Other Basque N-N sequences that are not compounds

When describing compounds, the Basque literature always mentions two classes of N-N sequences that have features outside the Xº level. First, N_1-N_2 appositive sequences of the kind *Saizarbitoria idazlea* lit. 'Saizarbitoria writer-the' usually denote a unique being within the class of N_2. However N_1 itself corresponds to the reading of the whole sequence. Therefore, neither a clearly syntactic nor a semantic head can be confirmed in appositive structures. These sequences necessarily take the attached article *-a*,[12] which is a syntactic-like feature.

The second kind of Basque non-dependency sequences traditionally taken as compounds are copulative. In such cases, N_1-N_2 indicates that the items in the class denoted by N_1 are added to the items in the class denoted by N_2. All nominal copulative constructions take both the article *–a* and the overt plural number mark *-k*: *senar-emazteak* lit. 'husband-wife-the'.

In short, the Basque bibliography mainly pays attention to the lack of a semantic/syntactic dependency relationship between constituents in N-N sequences that belong to different linguistic levels. So the dependency compound *itsasgizon* 'seaman' is shown in contrast with non-dependency sequences that usually are overt phrases.

2. Quantifying N-N sequences and applied linguistics

Solé (2002) sees the individualization or collectivization of non-distinguished individuals as a language universal that she explores from the point of view of Catalan collective nouns. For her part, Estopà (1990) discusses nouns involved in a complex determiner that somehow has a quantifying content of two types. First, she describes nouns that express a part, which are exemplified by *litre* 'litre', *mica* 'piece' and *quart* 'quart'. Second, she speaks about nouns that express a whole, which are exemplified by *grup* 'set', *quantitat* 'quantity' and *parella* 'couple'. Both authors seem to give a unified treatment to the field, if we conceptually and terminologically assume that there is a general kind of collective noun which is sometimes individualized (*una mica de pa* 'a piece of bread') and a general kind of collective noun which sometimes is used for collectivizing (count noun) individuals (*un grup de nois* 'a group' of boys'). Our informal label of 'quantifying nouns' corresponds to the individualizing nouns in the former class, and with the collectivization nouns in the latter. In fact, the quantified noun is usually a mass noun in the former and an individual (count) noun in the latter. All these nouns are involved in N-P-N sequences in Romance languages and English, but show a particular distribution between N-N sequences, measure phrases and N-P-N sequences in Basque.

---

[12] Regarding the obligatory use of the article *–a*, Odriozola (2006a) describes the different extralinguistic and linguistic uses of this suffix, which is clearly related to the productive neologism-creating process that the language has undergone in recent decades.

We are in great debt to these two works on several scores. First, we coincide with them concerning the semantic/lexical/syntactic field of collective nouns. Second, we take from these works both the terminology and the universal semantic concepts that may be relevant for distinguishing Basque quantifying N-N sequences from any other N-N sequence. In fact, the description of quantifying N-N sequences of a type that is not available in Romance languages (nor in English) is one of the basic contributions of this paper. Finally, as an additional contribution, we hope to further the applied linguistics aspects that Solés and Estopàs had in mind.

Section 2.1 provides morphological, lexical and syntactic data that can be useful for any kind of language-processing activity in fields where quantifying N-N sequences are frequently found. Section 2.2 is concerned with Basque spelling conventions for quantifying N-N sequences in language for special purposes.

2.1. Quantifying nouns and language processing

This section is meant to help in fields of applied linguistics such as automatic information recovery, reading recognition of quantifying nouns not listed previously, improvement of lexical information in (electronic) dictionaries, and language processing in general. Our aim is to provide the features of quantified and quantifying noun sequences that could be used in works of this kind. This paper cannot be concerned with the necessary formal mechanisms, but will provide information at three levels that presupposes language processing types using various kinds of prior linguistic information. First, we assume that all language processing in this field will be provided with syntactic information about the inflectional morphology and grammatical categories of the items concerned. Second, information on the lexical morphology of derivational processes will usually be available. Third, previous lexical information of three types might also be available. First, the possibility of detecting certain verbs can be very useful. Second, the accurate subcategorization of these and certain other verbs would give still more help. Third, previous subcategorical information about mass nouns would obviously be the most useful help of all. We will distinguish between part quantifying nouns of mass collectives (*ogi zati* lit. 'bread piece') and collective quantifying nouns of countable parts (*mutil multzo* lit 'boy set'). We follow Estopá (1996), Solé (2002), and Lorente (2002: §8.1.2.3) and distinguish between non-specific and specific quantifying nouns, in a way that is similar to the distinction between weak and strong quantifiers.

2.1.1. Part nouns

The features given by Solé that are relevant to our paper are the following: Nouns involved in Basque quantifying N-N sequences that denote a relation where the parts are not structured and can be separated from the whole. These parts do not have a specific function.[13]

[13] Part nouns that do not fulfil (some of) these conditions do not instantiate quantifications, although they may express meronymic relationships: *saguzar-hego* lit. 'bat wing', *mendi tontor* lit. 'moun-

A mass noun and a verb with a plural mark may be used as an indicator for part nouns, if the corresponding lexical and inflectional information is available.

2.1.1.1. Non-specific part nouns

The general distribution of these nouns in Basque syntax is found in quantifying N-N sequences and N-P-N sequences.

2.1.1.1.1. Nouns like *zati* 'piece' express a part of a mass (*bi ogi zati* lit. 'two bread piece').[14]

2.1.1.1.2. As described in section 1.2.2, some part nouns like *pitin* 'bit' need to appear with *bat* 'one/a', which should be expressed like this in dictionaries and in every (automatic) data-base.[15] As noted in section 1.2.1, they do not accept genitive phrases to the left of the N-N sequence.

2.1.1.1.3. Some general language nouns such as *ale* 'grain', *bola* 'ball', *izpi* 'ray' and *tanta* 'drop' pattern syntactically with *zati* but they express a particular meronymic relationship, since they denote the (smallest) natural form in which the mass appears. We assume that these nouns express non-specific quantification. New nouns of this type can be detected by means of lexical information about verbs such as *eratu* 'to form' in their intransitive counterparts.[16]

2.1.1.2. Specific part nouns

2.1.1.2.1. Specific part nouns with the measure phrase option. In Basque, quantifying nouns in this category do not accept postpositional phrases to the left of the N-N sequence. Furthermore, they may optionally appear in measure phrases that have a unique reading related to the quantification of the mass noun (Odriozola 2006b).

(7) a bi litro/botila esne/ardo
two litre/bottle milk/wine
'two litres/bottles of milk/wine'

b bi litro/kikara kafe
two litre/cup coffee
'two litres/cups of coffee'

---

tain top', *liburu atal* lit. 'book section', *esnegain postre* lit.'cream dessert'. See Odriozola (2004) for sequences like *lotsagabe* lit. 'shame lacking' (shameless).

[14] N-N sequences like *eroste unitate* lit. 'buying unit', *ur lagin* lit. 'water sample', *segundo frakzio* lit. 'second fraction' (fraction of a second) are found in specialized texts. Sometimes nouns of the general language take on specialized meanings in certain Basque quantifying N-N sequences in specialized texts: *zeramida zati* lit. 'ceramide piece'. On the other hand, *kantitate* 'quantity' expresses a special meronymic relation at all levels of the language: *zor kantitate* lit. 'debt quantity'. Sometimes quantifying nouns not previously listed impose a still more particular meronymic relation: *xurgapen koefiziente* lit. 'absorption coefficient', *inbertsio tasa* lit. 'inversion rate', *kontzentrazio indize* 'concentration index'.

[15] This problem of bordering between (lexical) quantifying nouns and (functional) complex quantifiers —i.e., the issue of the two optional readings of these nouns— is very common in human languages. Quantifying nouns in both English and Romance languages are usually involved in N-P-N sequences. Certain Georgian quantifying nouns may appear with quantified nouns case-marked genitive or absolutive.

[16] One of the few English counterparts of Basque quantifying sequences is that of the *drop* class: *water drop* is grammatical.

The Basque measure phrases in (7-9b) are a parametrical feature of this language[17] that is not as good an indicator in language processing, even when additional information about mass nouns is available. See the set of samples in (8-9).

(8) a Bi esne-behi ikusi ditut
Two milk cow seen I-have-them
lit. 'I have seen two milk cows'

b Bi behi-esne ikusi ditut
Two cow milk seen I-have-them
lit. 'I have seen two (types of) cow milks'

(9) a Bi esne-botila edan ditut
two milk bottle drunk I-have-them
'I have drunk two bottles of milk'

b Bi botila esne edan ditut
two bottle milk drunk I-have-them
'I have drunk two bottles of milk'

The different structures in (8-9) can only be distinguished using a combination of lexical information and subcategory information about nouns and verbs, inflectional information about verbs and syntactical information about word order. In fact, (8a-b) are N-N compounds, (9a) is a quantifying N-N sequence and (9b) is a measure phrase.

Let us now see in more detail the set of nouns that yield Basque measure phrases.

2.1.1.2.1.1. Unit nouns. Volume and weight unit nouns typically appear in measure phrases of both general Basque and language for special purposes. They are not totally standard as heads of quantifying N-N sequences like that in (9a), but they are not excluded. Languages for special purposes have automatically incorporated measure phrases not available in the general language. *Mol* expresses a weight that is specific for every chemical element. *Molekula* 'molecule' expresses the smallest part in which an element can appear, but also corresponds to a given weight. Of course, these concepts are scientifically determined and learned outside the general language in systematic studies. All the subcategories and language levels are distinguished by speakers. *Litro* 'litre' is a unit noun of both general and specialized level that tends to appear only in measure phrases (10a). Both *molekula* and *tanta* 'drop' (see section 2.1.1.1.3) denote the smallest part of a mass at a given knowledge level. However the speaker distinguishes the specificity of *molekula* and uses it both in measure phrases (10a) and in quantifying N-N sequences (10c), whereas the non-specific noun *tanta* is avoided in measure phrases (10b). Differences are even distinguished between the two nouns *molekula* and *mol* outside the general language in that *mol* tends to be used only in measure phrases (10a). We therefore conclude that Basque distinguishes between units, scientifically-determined smallest parts and smallest parts at the general level of the language.

(10) a Bi molekula/mol/litro sulfuriko
two molecule/litre sulphur

b *bi tanta sulfuriko
two drop sulphur

c Bi sulfuriko molekula
two sulphur molecule

d bi sulfuriko tanta
two sulphur drop

Units nouns are involved in a third construction of the type of *bi litroko bolumena* lit. 'two litre-of volume' that features both a postposition and a magnitude noun.

[17] Unlike English and Romance languages, German and Georgian have measure phrase of the Basque type.

This construction is available for every unit noun and every magnitude noun at all levels of the language. Lexical information about magnitude nouns combined with information about *-ko* and information about determiners could help in detecting any unit noun not previously listed in the bibliography.

2.1.1.2.1.2. Container nouns. Importantly, container nouns like *botila* 'bottle' and *koilara* 'spoon' typically appear in measure phrases, which shows that they are treated somewhat as units. However, they pattern with nouns like *drop* in that they can head quantifying N-N sequences without constraint. Therefore they fulfil the general condition of quantifying N-N sequences in that a quantifying reading of the left element is allowed, but they also allow a container/non-quantifying reading.[18] The two optional readings of container nouns contrast with the tendency toward a unique reading of all the other part nouns and with the single quantifying reading of some of them.

Basque measure phrases allow another set of nouns derived from container nouns such as *goilarakada* 'spoonful' that only take a quantifying reading even in N-N sequences, constrasting with *goilara,* which may take both. In any case, these derived nouns have a particular morphological feature that can be used as an indicator for the automatic detection of these constructions.

2.1.1.2.2. Specific part nouns without the measure phrase option. Basque has partitive numerals that are derived from the standard numerals and that appear as nouns in dictionaries. They are not allowed in measure phrases, but their particular morphology[19] may be used as an indicator.

(11) a bi opil hiruren
two muffin third
'two thirds of the/a muffin'

b bi opil hirurenak
two muffin third -DET-PL
lit. 'the two third of the muffin'

2.1.2. Collective nouns

Following the specifications made by Solé (2002) for Catalan, we shall also examine collective nouns in Basque that do not lexically express a specific function and that are made up of equal units that also have no specific function.[20]

Such nouns and quantifying N-N sequences take a verb in the singular, but they accept modifiers such as *aho batez* 'unanimously' and *batera* 'altogether' at least when the part noun corresponds to a living being —a fact that could be a good indicator in language processing. Moreover, verbs such as *bildu* 'to gather', *elkartu* 'to

---

[18] (i) a. Bi garagardo-botila edan nituen
two beer bottle drunk I-had-them

b. Bi garagardo-botila apurtu nituen
two beer bottle broke I-had-them

English has this type of N-N sequences, i.e. *beer bottle*, although following Castillo (2001) their reading never is a quantifying one. That is, they are not quantifying N-N sequences, in the informal sense used in this paper.

[19] As in English and Romance languages, 'half' is expressed in Basque by a noun that is not derived.

[20] Therefore we are not concerned with nouns like *batzorde* 'commission', *bilera* 'meeting', *abesbatza* 'chorus', which will be assumed not to instantiate any quantification, although they are collective nouns that may head Basque N-N sequences.

join/meet', *banandu* 'to separate', *desegin* 'to split', *sakabanatu* 'to spread', can also be used.

We will distinguish again between specific nouns and non-specific nouns.

2.1.2.1. Non-specific nouns

Nouns like *multzo* 'group' are able to appear in both quantifying N-N sequences and following genitive PPs that are identical to those that appear with part nouns.[21] Some nouns like *pilo* 'lot' do not allow a genitive PP, as noted before. Some collective nouns like *talde* 'set/team' accept two kinds of N-N sequences, i.e. *mutil talde* lit. 'boy set/group' and *futbol talde* 'football team' or *marketin talde* lit.'marketing group'. With *bat* 'one/a' the former can instantiate a quantification whereas *futbol talde* (which has a grammatical counterpart in English) is a N-N compound headed by a fully nominal element. In fact, the former shows a meronymic relationship, whereas the latter takes a standard hyponymic reading with respect to the head *talde* 'team'.

2.1.2.2. Specific nouns

Basque specific collective nouns involved in quantifying N-N sequences are derived from numerals: *bikote* 'pair' (cif. *Bi* 'two'), *hirukote* 'trio' (cif. *hiru* 'three'), *laukote* 'quartet (cif. *lau* 'four'). These nouns do not appear after genitive PPs.

2.2. Quantifying nouns and language for special purposes

The lexicographical bibliography of Basque has dealt at length with the issue of spelling rules for Basque compounds, for two very different reasons. On the one hand, Basque speakers can create N-N sequences spontaneously in language performance. Let us remember that English N-N sequences are taken either as phrases or as compounds, and that new N-N sequences are created again and again. It is the same for Basque N-N sequences, but all the bibliography preceding this paper coincides in stating that such constructions are compounds, regardless of their (total) lack of stability in the lexicon.

In section 1.2, we showed that besides quantifying N-N sequences, Basque has among other N-N sequences dependency compounds and appositive constructions. Let us consider a simple dependency compound of the general language, like *hiri zubi* lit. 'city bridge'. Standard spelling rules allow both a hyphen (*hiri-zubi*) and writing the two nouns separately (*hiri zubi*). In fact, given the meaning of these two general language nouns, Basque readers do not hesitate to decode the new sequence *hiri zubi* as a dependency compound. However, let us now see two examples of language for special purposes, where word formation has been very intense in recent decades:

---

[21] *Kopuru* 'quantity' expresses a special meronymic relationhisp at all levels of the language with items that are taken as count nouns in Basque: *diru kopurua* lit. 'money quantity-the', *pertsona kopurua* lit. 'person quantity-the'.

(18) a hidrogeno-zubi
hydrogen bridge
lit. 'hydrogen bridge'

b disulfuro zubia
disulphur bridge-DET
lit. 'disulphur bridge'

The sentences in (18a-b) can only be decoded correctly with the help of specialized (extralinguistic) knowledge.[22] Needless to say, one must know the meanings of 'hydrogen' and 'disulphur', but even this is not enough, given the fact that that some speakers with a knowledge of chemistry do not have sufficient linguistic awareness to distinguish whether a dependency or an appositive relation is involved in (18a-b). The fact is that *hidrogeno zubi* is a kind of bridge or bonding in which hydrogen takes part, whereas *disulfuro zubia* is a bridge called *disulfuro*. That is, the former is a dependency compound (section 1.1.2), and the latter is an appositive construction (section 1.2.2). Since this difference must be decoded, Basque science writers systematically use the hyphen as a convention to (orthographically) distinguish dependency compounds like (18a) from appositive constructions like (18b), which are written separately at all levels of the language. Not all Basque speakers need to understand the importance of decoding (18a) and (18b) through different means, since speakers do not have any problem in making and understanding constructions like (18a-b) in the general language, but spelling conventions are clearly a help in specialized writing.

However, this decision about using the hyphen in dependency compounds in specialised language has been extended to quantifying N-N sequences, in order to distinguish the latter from the former. This might seem trivial but it is not. The efforts of specialised writers to distinguish (18a) from (18b), or to distinguish a hyponymic relation from a meronymic have served not only to clarify Basque spelling, but have also helped to heighten awareness of Basque morphology, while at the same time furthering special knowledge of (in this case) Chemistry. Therefore, both kinds of knowledge would eventually be improved if writers and readers were aware of the meronymic relationship expressed by a N-N sequence like, for instance, *hidrogeno kantitate* lit. 'hydrogen quantity'. The hyphen is closely associated with dependency relationships in Basque, but such a relationship does not exist in quantifying N-N sequences and therefore the hyphen should not be used. Indeed, hyphens must not be used unless they serve a purpose. Moreover, there are grammatical reasons for not using the hyphen in quantifying N-N sequences. These structures are easily decoded by readers and writers at all levels of the language, either because of the clear thematic relationship denoted by the quantifying noun, or because of the clear quantifying nature of the head. The concept *hidrogeno* needs to be learn in Chemistry, but the relationship that any noun has with *kantitate* does not. And finally, all the quantifying nouns not previously listed but detectable by the methodology outlined in 2.1. are now being used correctly at all levels of the language.[23] In short, quantifying

[22] It should be remarked that in these examples, orthography options are related to existing/non-existing realities. In fact, the problems of ambiguity in Basque N-N sequences are concerned with cases where two existing realities compete for understanding.

[23] For instance, the hyponymic sequences *turismo-elementu* lit. 'tourism element' or *inportazio-prezio* lit. 'importation price' can be compared to the meronymic sequences *turismo kantitate* lit. 'tourism quantity', *importazio tasa* lit. 'importation rate'. There is no doubt that the latter are easily decoded, whereas the former are at least not familiar in the sense of Bauer (2001: §3.2).

N-N sequences are not based on unexpected relationships, but on relations previously included in the reader's linguistic knowledge. Knowing this fact may also help in following the strict spelling conventions that language for special purposes needs. In short, the tendency of today's (specialized) Basque is to write *hidrogeno-zubia* 'hydrogen bridge' (dependency compound), *hidrogeno kantitate* (quantifying N-N sequence) and *disulfuro zubi* (appositive N-N sequence).

References

Artiagoitia, X., 2002, "The functional structure of the Basque noun phrase", in X. Artiagoitia, P. Goenaga and J. A. Lakarra (eds.), *Erramu Boneta*: *Festschrift for Rudolf P. G. de Rijk,* Supplements of *ASJU*, 73-90.

Azkarate, M., 1990, *Hitz elkartuak euskaraz*, Mundaiz, Donostia.

Bauer, L., 2001, "Morphological productivity"*, Cambridge Studies in Linguistics* 95. Cambridge U. P.

Becker, T., 1993, "Compounding in German", *Rivista di Lingüística* 4: 1, 5-36.

Borer, H., 2005, *In Name Only,* Oxford U. P., New York.

Buenafuente, C., 2007, *Lexicalización en la formación de compuestos en español*, Doctoral Dissertation, Universitat Autònoma de Barcelona.

Castillo, J. C., 2001, *Thematic Relations between Nouns,* Doctoral Dissertation, University of Maryland.

Estopà, R.. 1996. "Noms que formen part d"un determinant complex", *Series monografiques*, Universitat Pompeu Fabra.

Etxeberria, U., 2005, *Quantification and Domain Restriction in Basque*, Euskal Herriko Unibertsitatea.

Euskaltzaindia, 1987, *Hitz-elkarketa/1. LEF batzordearen lanak,* Bilbao.

—, 1991, *Hitz-elkarketa/3*, Bilbao.

Gràcia, Ll., 1994, *Estructura argumental i herència en morfologia*, Universitat de Girona.

—, et al., 2000, *Configuración morfológica y estructura argumental: léxico y diccionario*, Servicio Editorial de la Universidad del País Vasco.

Grimshaw, J,. 1990, *Argument Structure*, The MIT Press, Cambridge.

Haspelmath, M., 2002, *Understanding Morphology*, Arnold-Oxford U. P. Inc., New York.

Lieber, R., 1993, "Compounding in English", *Rivista di Lingüística* 4: 1, 79-96.

Lorente, M., 2002, "Altres elements lèxics", in J. Solà (dir.), *Gramàtica del català contemporani I: 8*, Empúries, Barcelona, 831-888.

Marchand, H., 1960, *The Categories and Types of Present-day English Formation. A Synchronic-Diachronic Approach,* Otto Harrassowitz-Wiesbaden.

Odriozola, J. C. & E. Pérez Gaztelu, 2002, "Aposizioa euskal hitz-elkarteetan", in Artiagoitia, Goenaga & Lakarra (eds.), 467-478.

—, 2004, "Construcciones con *gabe* "sin" en vasco", in Pérez Gaztelu et al. (eds.), 355-392.

—, 2006a, "(Basque) natural phrases for artificial languages". *Andolin Gogoan: Essays in Honour of Prof Eguzkitza,* UPV/EHU, Bilbao, 707-724.

—, 2006b, "Measure phrases in Basque", In J. A. Lakarra & J. I. Hualde (eds.), *Studies in Basque and Historical Linguistics in Memory of R.L. Trask. ASJU* 40:1-2: 739-762.

—, 2007, "*Bilbo-Behobia* bezalako tandem-elkarteak: hiru osagai baina morfologi prozesu bakarra", *Iker* (19) (in press).

Pérez-Gaztelu, E., 2004, "Tipos de compuestos", in Pérez-Gaztelu et al., 109-162.
—, I. Zabala, Ll. Gràcia (arg.), 2004, *Las fronteras de la composición en lenguas románicas y en vasco,* Universidad de Deusto, San Sebastián.
Pustejovsky, J., 1998, *The Generative Lexicon*, The MIT Press, Cambridge.
Rainer, F. & Varela, S., 1992, "Compounding in Spanish", *Rivista di Linguistica* 4: 1, 117-142.
Solé, E., 2002, *Els noms col.lectius Catalans. Descripció I reconeixement*, Doctoral Dissertation, Unibersitat Pompeu Fabra.
Suñer, A., 2004, "Los procesos de lexicalización", in Pérez-Gaztelu et al. (eds.), *Las fronteras de la composición en lenguas románicas y en vasco*, Deustuko Unibertsitatea, Bilbao.
Trask, R. L., 2003, "The Noun Phrase: nouns, determiners and modifiers; pronouns and names", in J. I. Hualde & J. Ortiz de Urbina (eds.) *A Grammar of Basque*, Mouton de Gruyter, Berlin-New York.