Light Nouns and Term Creation in Basque

Igone Zabala
Dept. of Basque Language and Communication
University of the Basque Country UPV/EHU
igone.zabala@ehu.eus

Igone Zabala és doctora en Filologia Basca (1993) per la Universitat del País Basc (UPV/EHU). Tot i que la seva tesi doctoral i treballs posteriors s’emmarquen en l’àmbit de la sintaxi i de la morfologia lèxica, fa més de deu anys que la seva recerca se centra en el discurs especialitzat i la terminologia, amb la mirada centrada en el desenvolupament lèxic i estilístic de l’euscar en contextos especialitzats. Actualment és membre del grup de recerca IXA, especialitzat en procesament del llenguatge natural. És professora de Comunicació especialitzada en euscar a la Facultat de Ciència i Tecnologia i a la Facultat de Medicina i Infermeria de la Universitat del País Basc (UPV/EHU). És també la responsable del programa Terminología Sareak Ehunduz (TSE) ‘teixint xarxes terminològiques’, promogut pel Vicerectorat d’Euscar de la UPV/EHU.

Resum

Noms lleugers i creació de termes en l’eusquera

Aquest article estudia l’activació de valors especialitzats en un grup de noms lleugers (NL) de l’eusquera del tipus talde, ‘grup’, des del punt de vista de l’escala de referencialitat (ER) (Simone i Masini, 2009 i 2014). L’estudi demostra que els processos de terminologització dels NL es produeixen en el sentit descrit per Adelstein (2004 i 2007) per als ultranoms, i postula que els paratermes poden ser caracteritzats com a «noms de suport especialitzats» situats en l’ER en una posició més baixa que la dels termes.

Paraules clau: noms lleugers; creació de termes; paratermes

Abstract

This paper investigates the activation of specialized values in a set of Basque Light Nouns (LNs) such as talde ‘group’, from the viewpoint of the referentiality scale (RS) (Simone & Masini, 2009, 2014). We demonstrate that the terminologization of LNs happens in the sense described by Adelstein (2004, 2007) for ultranouns, and we postulate that paraterms can be characterized as “specialized support nouns” positioned lower than terms in the RE.

Keywords: light nouns; term creation; paraterms
1 Introduction

This paper focuses on a set of Basque nouns, such as talde ‘group’, zati ‘part’ and mota ‘type’, which occur prototypically as N2 in a set of root compounds (N1LN2) called mugakizun bereziko izen elkarteak ‘compound nouns with special base’ in Basque literature (Azkarate, 1990; Azkarate & Perez Gaztelu, 2014): ikasle talde ‘group of students’, gazta zati ‘piece of cheese’, ardo mota ‘wine type’. These nouns can be characterized as Light Nouns (LNs) (Simone & Masini, 2009, 2014). According to Simone & Masini the distinguishing characteristic of LNs – which in Romance languages and English often occur as N1 in binominal constructions of the N1 of N2 type (bit of bread, lot of friends, sorta di indifferenza ‘sort of indifference’, mancia-ta di riso ‘fistful of rice’) – is that they are lower in referential force (RF) than full-fledged lexical nouns. Due to their referentiality weakness, LNs are semantically unsaturated and consequently dependent on another element in order to complete their reference (N2 in Romance languages or English and N1 in Basque). Moreover, Simone & Masini (2014) postulate that languages have Referential Force Modulators (RFM), namely tools that increase or decrease the RF according to discourse needs. They demonstrate that under certain syntactic conditions the referentiality of some LNs weakens, which is reminiscent of the behavior of “light verbs”: LNs can modulate their RF tendency to lose their lexical contain and acquiring a more grammatical sense that serves to modify N2.

Basque LNs such as talde ‘group’ have been exhaustively described by Azkarate & Perez Gaztelu (2014). These nouns present subclasses analogues to those studied by Simone and Masini (2009, 2014): Group and Partitive Quantifiers, Taxonomic Nouns, Approximators, and Support Nouns. Our motivation for studying these nouns is twofold. On the one hand, there is an applied motivation, since LNs frequently cause a lot of noise in automatic terminology extraction (Estopó & Lorente, 2008, 2010). According to Estopó and Lorente the abovementioned noise is due to the high presence of these nouns either in general or specialized corpora, and to the fact that they can activate either general or specialized values in those corpora. On the other hand, there is a theoretical interest in the manner in which these nouns, characterized by a low RF, can activate different specialized values. This theoretical interest is supported by one of the foundational tenets of the Communicative Theory of Terminology (CTT) (Cabré, 1999, 2001). According to the CTT, terms and words are activations of different semantic features of a lexical item that take place in different contexts of use. The assumption that terms are part of natural language’s lexicon blurs the boundaries between words and terms, which were considered elements of different semiotic systems in the General Theory of Terminology (GTT) of Wüster (1979), and consequently, this makes an integrated representation of specialized and non-specialized values of lexical items necessary.

Within the CTT framework, Adelstein (2004, 2007) postulates that the specificity of the specialized lexicon lies in the imbricated semantic configuration of meanings, determined by several factors of the communicative situation, by sentence factors and by syntactic-semantic factors. In order to demonstrate her hypothesis, Adelstein analyzed the semantic configuration of some lexical items, such as madre ‘mother’, which present several specialized and non-specialized values, following the polysemic model of Pustejovsky (1995) called the Generative Lexicon. The Generative Lexicon model postulates that lexical meanings in context are generated via activation of a part of the elements that make up the semantic structure of lexical items. Semantic structure is described by compositional elements with predicative nature. Using an adaptation of Pustejovsky’s model, Adelstein concludes that specialized meanings are characterized by two basic properties: higher referential specificity and greater imbrication of semantic information. One of the aims of this paper is specifically to analyze if Adelstein’s conclusions can be extrapolated to Basque LNs that have activated specialized values.

Section 2 is devoted to a lexical-semantic characterization of Basque LNs in general usages: morphosyntactic tests are used in order to establish a RF scale for these nouns. Section 3 analyzes specialized values activated by Basque LNs, and these specialized values are compared with general uses, in order to check Adelstein’s postulates about the activation of specialized values of lexical items. Section 4 is devoted to the conclusions.

2 Basque LNs in general contexts

Azkarate (1990) and Azkarate & Perez Gaztelu (2014) carry out a lexico-semantic characterization of LNs that occur as the second member of root compounds of the type N1LN2. These compound nouns are called mugakizun bereziko izen elkarteak ‘compound nouns with special base’ (CNSB) in Basque literature. Even though these authors distinguish seven subtypes of CNSB, we will limit our study to three subtypes, since they are easily comparable with the subclasses investigated by Simone & Masini (2009, 2014): a) multzo-elkartek ‘group compounds’ (ikasle talde lit. student group ‘group of students’), b) zati-elkartek ‘part compounds’ (gazta zati lit. cheese piece ‘piece of cheese’), and c) mota-elkartek ‘type compounds’ (ardo mota lit. wine type). In these compounds N1 is a full-fledged noun, whereas N2 (talde ‘group’, zati ‘part’, mota ‘type’) is an LN that takes on a grammatical meaning projected onto N1. These LNs are analogous to Group Quantifiers,
Part Quantifiers, and Taxonomic Nouns in Simone & Masini (2014). Table 1 summarizes LNs belonging to the three types examined in this paper following the classification in Azkarate & Perez Gaztelu (2014).

Among the nouns in Table 1, those belonging to the multzo ‘group’ and zati ‘part’ types can be classified in the semantic group of Quantifier Nouns. The former are denominated by Bosque (1999, 18-22) as Sustantivos de Grupo ‘group nouns’, and the latter, Sustantivos Acotadores ‘encloser nouns’. Simone & Masini (2014) refer to the zati ‘part’ type nouns as Partitive Quantifiers, and to the mota ‘type’ nouns as Taxonomic Nouns. In Table 1 we have adapted and integrated the terminology used by these three authors.

Simone & Masini (2014) use an array of tests in order to measure the referentiality of LNs used in N1 de/of N2 constructions in Romance languages and English: semantic headness, internal and external agreement, and constituency. Some of these tests can be used in order to characterize compounds with LNs in Basque (N1LN2) and to compare them with full-fledged lexical root compounds.

First of all, the ‘IS A’ condition of Allen (1978) demonstrates that in root compounds with referential nouns N2 is the formal and semantic head of the compound, since the compound’s reference is the same as N2 (1).

<table>
<thead>
<tr>
<th>LN2 = Sustantivos de Grupo ‘group nouns’ (Bosque 1999)</th>
<th>LN2 = Partitive Quantifiers (Simone &amp; Masini 2014) or Sustantivos Acotadores ‘encloser nouns’ (Bosque 1999)</th>
<th>LN2 = Taxonomic Nouns (Simone &amp; Masini 2014)</th>
</tr>
</thead>
<tbody>
<tr>
<td>aldra ‘group, horde, crowd’</td>
<td>atal ‘part, fragment, section’</td>
<td>era ‘manner, class, kind’</td>
</tr>
<tr>
<td>osto ‘group, crowd, army’</td>
<td>hondar ‘rest, remainder, residue’</td>
<td>gisa ‘manner, way, fashion’</td>
</tr>
<tr>
<td>saldo ‘multitude, crowd, army’</td>
<td>klase ‘class, category, standing, status’</td>
<td></td>
</tr>
<tr>
<td>talde ‘group, team’</td>
<td>izpi ‘blade, wipe, filament, strand, ray, beam’</td>
<td></td>
</tr>
<tr>
<td>tanda ‘group’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>tropel ‘horde, crowd’</td>
<td>koskor/kosko ‘piece, chunk, bit, end’</td>
<td>manera ‘manner, way’</td>
</tr>
<tr>
<td>mordo ‘big group, flock’</td>
<td>mutur ‘tip, end, piece’</td>
<td>modu ‘manner, way’</td>
</tr>
<tr>
<td>andana ‘bunch, cluster, big amount, big quantity’</td>
<td>parte ‘part, piece, section’</td>
<td></td>
</tr>
<tr>
<td>mordoxka ‘appreciable quantity or amount’</td>
<td>punta ‘tip, point, end’</td>
<td>molde ‘way, manner, appearance, shape’</td>
</tr>
<tr>
<td>mulko ‘bunch, cluster, big amount or quantity’</td>
<td>puska / pusketa ‘piece, bit, chunk’</td>
<td></td>
</tr>
<tr>
<td>multzo ‘group, set’</td>
<td>zati ‘part, piece, segment’</td>
<td>mota ‘type, class, kind, sort’</td>
</tr>
<tr>
<td>pare ‘pair, couple’</td>
<td>apur ‘few, pinch, not much’</td>
<td></td>
</tr>
<tr>
<td>pila / pilo ‘pile, stack, big amount’</td>
<td>pittin / pixka / poxi ‘few, pinch, not much’</td>
<td>suerte ‘type, class, kind, sort’</td>
</tr>
<tr>
<td>sail ‘group, department, section’</td>
<td>puska ‘large, big, excessive’</td>
<td></td>
</tr>
<tr>
<td>meta ‘lot, pile’</td>
<td>tanta / tanto ‘drop, bit, speck’</td>
<td>tailu ‘look, appearance’</td>
</tr>
<tr>
<td>sorta ‘bunch, handful, bouquet, string, strand’</td>
<td>xorta ‘drop, bit, speck’</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: LNs that take part in Basque N1LN2 mugakizun bereziko izen elkarteak ‘compound nouns with special base’ according to Azkarate & Perez Gaztelu (2014)
(1) a. behi-esne IS esne (N2) 
   cow-milk milk
   ‘dairy milk’

   b. esne-behi IS behi (N2) 
   milk-cow cow
   ‘dairy cow’

However, when the second member of the compound is an LN, it does not work as a semantic head, but rather as some kind of modifier (2).

(2) a. landare mota bat IS landare (N1) (Taxonomic Noun) 
   plant type a plant
   ‘a type of plant’

   b. ogi zati bat IS ogi (N1) (Partitive Quantifier) 
   bread piece a bread
   ‘a piece of bread’

   c. ikasle talde bat ARE ikasleak (N1) (Group Quantifier) 
   student group a students
   ‘a group of students’

The fact that LN2 is not the semantic head of the construction is more evident when we replace the compound with N1, since LN2 can be dropped without any substantial change in the meaning (4). On the contrary, the replacement gives rise to a completely different meaning in root compounds with full referential nouns (3).

(3) a. Behi-esnea edan dut. 
   cow-milk-DET drink AUX-3sg-1sg
   ‘I drank dairy milk.’

   b. *Behia edan dut 
   cow-DET drink AUX-3sg-1sg
   ‘I drank the cow.’

   c. Esne mota berri bat edan dut. 
   milk type new a drink AUX-3sg-1sg
   ‘I drank a new type of milk.’

Furthermore, in root compounds with referential nouns, N1 acquires a generic meaning and any kind of modification is disallowed (5b). On the contrary, in LN constructions several kinds of modifiers related to N1 are allowed (6b).

(5) a. behi-okela 
   cow-meat
   ‘cow meat’

   b. *behi zahar okela 
   cow old meat
   ‘old cow meat’

(6) a. gazta puxka 
   cheese piece
   ‘piece of cheese’

   b. gazta zahar puxka 
   cheese old piece
   ‘piece of old cheese’

Constituency can be checked using an array of tests aimed at demonstrating the ability to take part in referential chains. Whereas in root compounds with a high RF N1 cannot be part of reference chains (7b), both members of LN compounds can participate in reference chains (8).

2.1 Weakened LNs

There are some contexts in which LNs present even weaker RF, up to the point of behaving as degree modifiers. These quantifier LNs allow either singular or plural agreement with the verb (9), and as well as LN2 in compounds, they can take part in measure phrases as LN1 (9b, c), acquiring a more grammatical value. Consequently, we can conclude that LNs in measure phrases are positioned lower in the nouniness scale (Simone & Masini, 2014) than LNs in compound nouns.
Light Nouns and Term Creation in Basque
Igone Zabala

(9) a. Praka pila (bat) dau(z)kat. (Degree modifier NL) trousers pile (a) have-3sgE/(3plA)-1sgE ‘I have a lot of trousers.’
b. Pila bat praka dau(z)kat. (Measure phrase) pile a trousers have-3sgE/(3plA)-1sgE ‘I have a lot of trousers.’
c. Dezma bat praka dau(z)kat. (Measure phrase) dozen a trousers have-3sgE/(3plA)-1sgE ‘I have a dozen trousers.’

In contrast with the LN in (8a), the noun pila ‘pile, stack’ is not able to take part in referential chains, which demonstrates its referential weakness (10a).

Another case of LNs with a weakened RF is the use of Taxonomic Nouns (11) as Approximators (12).

(10) a. *Praka pila bati daukat, baina proi ez zait gustatzen. trousers pile ai have-3sgA-1sgE but proi no AUX-1sgA-1sgD like ‘I have a pile of trousers, but I don’t like it (=the pile)’
b. Prakai pila bat dau(z)kat, baina proi ez zaizkit gustatzen. trousersi pile a have-3sg(/3plA)A-1sgE but proi no aux-1plA-1sgD like ‘I have a pile of trousers, but I don’t like them (=the trousers).’

(11) a. Gaitz mota hori sendatzeko tratamendu a (Taxonomic noun) disease type that cure-to treatment ‘Treatment to cure that type of disease’
b. Gaitz mota hori sendatzeko tratamendua (Approximator) disease type that cure-to treatment ‘Treatment to cure that type of disease’

(12) a. Baihun, zer gaixo klasa zara zu? (Approximator) but what sick-person kind are you? ‘But, what kind of sick person are you?’
b. Esperanto-modu bat hitz egin zuen. (Approximator) Esperanto-manner a word do AUX-past-3sgE-1sgA ‘He spoke a sort of Esperanto.’
c. Kontu pare bat esan behar di(zki)zuet. thing couple a tell must AUX-3sgA/(3plA)-2plD-1sgE ‘I must tell you a couple of things.’

2.2 Strengthened LNs

Simone & Masini (2014) only study the weakening of LNs; however, there are also several contexts in which LNs are strengthened. For example, the Basque noun talde ‘group’ often works as a number quantifier (13a) and can be replaced with the indefinite quantifier batzuk ‘some’ without changing its meaning significantly (13b). However, it also gives rise to the lexicalized compound gizatalde ‘human group’, which is a true collective noun (14a). In this case, the noun talde ‘group’ can replace the whole compound without significant meaning change (14b), but it cannot be replaced with the quantifier batzuk ‘some’ (14c), which evidences its strong referentiality.

(13) a. Gizon talde bat dago kalean. man group a is street-deT-in ‘There is a group of men in the street.’
b. Gizon batzuk daude kalean. man some street-deT-in ‘There are some men in the street.’

(14) a. gizatalde baten suntsipena men-group a -GEN extinction-deT ‘The extinction of a human group.’
b. talde osoaren suntsipena group whole-GEN extinction-deT ‘The extinction of the whole group.’
c. gizon batzen suntsipena man some-GEN extinction-deT ‘The extinction of some men.’

The following step in the strengthening of the noun talde ‘group’ takes place when the N1 of the compound does not refer to the constitutive element of the group. For instance, in (15a) the noun gizarte ‘society’ is a collective noun, and the constitutive element of N2 (talde ‘group’), which should be ‘person’, is saturated in the lexical level, becoming a “shadow argument” (Pustejovsky, 1995, 63-64). This value in the semantic configuration implies an increase in the RF of the noun talde ‘group’, which is evidenced by the possibility to replace the whole compound (15b), and by the differ-
ent reading that the phrase acquires when the whole compound is replaced with N1 (15c).

(15)

  a. gizarte-talde boteretsuak
      society-group powerful-DETpl
     ‘Powerful social groups’
  ~ b. talde boteretsuak
      group powerful-DETpl
     ‘Powerful groups’
  * c. gizarte boteretsuak
      society powerful-DETpl
     ‘Powerful societies’

The noun talde ‘group’ can be considered even semantically stronger in compound nouns that refer to sports teams. In these cases, in addition to the shadow argument saturated in the lexicon, which is always interpreted as ‘person’, the group is characterized by a purpose, giving rise to a more complex semantic configuration. Notice that when this semantic configuration happens in Spanish or in English a different noun is used: equipo and team respectively. In this case, we observe that the whole compound can be replaced with N2, but not with N1.

(16)

  a. Futbol-talde irabazlea
      football-group winner-DET
     ‘The winning football team’
  ~ b. Talde irabazlea
      group winner-DET
     ‘The winning team’
  * c. Futbol irabazlea
      football winner-DET
     ‘The winning football’

In short, some of the values of the nouns examined in this paper can be considered LNs. Furthermore, in accordance with the prediction by Simone & Masini (2014), the RF is both variable, since different noun classes are characterized by different levels of RF, and modulable for different discourse requirements. Simone & Masini (2014) only study the weakening of some LNs, which give rise to more grammatical and less nominal values. In this paper, besides some examples of weakening of Basque LNs, we have also presented some examples of strengthening, related to more complex semantic configurations.

3 Light Nouns with specialized values

The process of Basque-language standardization and revitalization started in the late twentieth century. Thanks to the recognition of its status as an official language, in the 1980s, Basque started to be used for specialized purposes, mainly in academic contexts. The change in the language’s status determined a massive creation of terminology, in order to satisfy the academic communication needs. With the aim of exploring the contribution of the LNs analyzed in this paper to the terminologization processes in Basque, which has been very active in recent decades, taking the nouns at Table 1 as our starting point, we searched in several dictionaries and terminology databases.

According to the information obtained in these searches, the LNs that have activated specialized values are those collected in Table 2.

From the analysis of the information in Table 2, we have obtained a set of significant conclusions concerning the activation of specialized values in Basque LNs.

• Weakened RF nouns (degree modifiers and approximators) do not usually activate specialized values, which is coherent with the greater referential specificity of specialized meanings (Adelstein, 2004, 2007).

<table>
<thead>
<tr>
<th>GROUP NOUNS</th>
<th>PARTITIVE NOUNS</th>
<th>TAXONOMIC NOUNS</th>
</tr>
</thead>
<tbody>
<tr>
<td>talde</td>
<td>atal</td>
<td>klase</td>
</tr>
<tr>
<td>multzo</td>
<td>hondar</td>
<td>Modo</td>
</tr>
<tr>
<td>tropel</td>
<td>izpi</td>
<td></td>
</tr>
<tr>
<td>mondo</td>
<td>mutur</td>
<td></td>
</tr>
<tr>
<td>sorta</td>
<td>punta</td>
<td></td>
</tr>
<tr>
<td>sail</td>
<td>zati</td>
<td></td>
</tr>
<tr>
<td></td>
<td>sail</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: LNs that have activated specialized values in Basque
• Some nouns (talde ‘group’, hondar ‘residue, remnant’, sorta ‘beam, sheet’) have activated different specialized values in several domains, whereas other nouns have only activated one specialized value in a particular domain (tropel ‘bunch’, mota ‘type’).
• Sometimes, the consequence of the activation of different specialized values in several areas is a great number of different equivalents in other languages. For instance, the Basque noun sorta has at least six equivalents in English (‘beam, sheet, shipment, set, bunch, bundle’) and at least four equivalents in Spanish (‘haz, remesa, pliego, juego’).
• The case of the noun sail must be emphasized, since it activates two kinds of values: some values with a sense of group (‘series, class, plantation’) and other values with a partitive sense (‘department, section, division, item’).
• Finally, in traditional uses the noun talde ‘group’ appears combined mainly with [+animate] nouns (abere talde ‘group of domestic animals’, adri talde ‘flock of sheep’, lagun talde ‘group of friends’, antxoa talde ‘anchovy shoal’). However, with the revitalization and modernization of Basque it has been acquiring more and more uses combining with [+animate] nouns, principally specialized uses (Azkarate & Gaztelu, 2014). This change has been interpreted as a mark of dependence on dominant languages such as English or Spanish. However, in our opinion, this tendency can also be related to the preponderance of [+animate] and [+human] discursive topics in general uses, as opposed to a greater thematic diversification in specialized uses.

In this section we try to demonstrate that specialized uses of LNs are positioned very high in the referentiality scale. Nevertheless, at least two levels can be distinguished in this scale. At the highest level are the values of nouns that have become true terms, such as talde ‘group’, multzo ‘set’, and sail ‘series’ in mathematics. Specialized values that can be considered “para-terms” (Estopà & Lorente, 2008, 2010), such as talde ‘group’ in finantza-talde ‘financial holding group’, odol-talde ‘blood group’ (Med.), or talde prostetiko ‘prosthetic group’ (Chem.), are located lower in the scale.

3.1 True terms

LNs that have given rise to true terms, such as talde ‘group’ in mathematics, have stopped being semantically unsaturated, and consequently, they are not dependent on other nouns in order to complete their reference. Therefore, they can be defined independently in a particular specialized domain. Furthermore, they take part in more complex terms, either combined with several kinds of adjectives (talde lineal ‘linear group’, talde abeldar ‘abelian group’, talde simetrikoa ‘symmetric group’, talde trukakor ‘commutative group’, talde ez-trukakor ‘noncommutative group’) or modifiers of other nouns (talde-teoria ‘group theory’). Conversely, paraterms cannot be defined independently, since they acquire their specialized value thanks to other elements with which they combine to give rise to complex terms.

In order to test the characterization of specialized values of lexical items by Adelstein (2004, 2007), we will compare the definition of the mathematical term talde ‘group’ in the encyclopedic dictionary of science and technology ZT Histegia (Elhuyar, 2009) with the definitions of general values in the normative dictionary of the Basque Language Academy Euskaltzaindiaren Histegia (Euskaltzaindia, 2012).

Definition 1: ZT Histegia (Elhuyar, 2009, 1268)
"talde: Mat. Multzo batez eta barne-eragiketa batez (batuketa) osatutako egitura matematikoa, non eragiteak propietate bakoizko betetzeko dituen: elkartze-legea, elementu neutroa badu \((a + o = o + a = a)\) eta elementu guztiek simetrikoa dute \((a + \text{elementu emanik}, a'\text{badago, non }a + a'\text{ = o den})\)." "group: Mat. mathematical structure that consists of a set and an internal operation (addition), in which the operation has the following properties: associative property, neutral element \((a + o = o + a = a)\), and every element has a symmetric element (given an element \(a\), there exists \(a'\), so that \(a + a'\ = o\))."

Definition 2: Euskaltzaindiaren Histegia (Euskaltzaindia, 2012, 747)

\textbf{talde} 1 iz. Mota bereko izaki bizidunen multzoa… ‘Set of living beings of the same type.’
2 iz. Kirol eta joko batzuetan, kopuru jakin bateko jokalari multzoa, beste multzo baten aukra jokatzen duena… ‘In some sports and games, a group with a specific number of players that competes with another team.’
3 iz. Gauzen multzoa… ‘Set of things.’

Here we should recall Adelstein’s claim that specialized meanings of lexical items are characterized, on the one hand, by a great referential specificity, consisting in a bigger predication number, a higher precision in the extension of semantic arguments, and a higher intensity of the referents of such arguments. On the other hand, the semantic configuration of specialized meanings should be more dense or complex, due to the existence of several imbrication levels between the activated predicates (more entailment relations).

General meanings of talde ‘group’ reflected in the dictionary of Euskaltzaindia (2012) (Definition 2) can be arranged according to their referential specificity as \(3 < 1 < 2\) in a scale from lowest to highest. In fact, in subentry 3 the semantic argument can be anything, and there is just one predication (‘to be constituted by several elements’). In contrast, in subentry 1, the precision in the extension of the semantic argument is higher, since the elements of the set are required to be ‘living beings’ and, additionally, the intensity of the argument is higher: elements must share some properties,
which should give rise to a second predication (‘to share properties’). The most complex semantic configuration among the three collected in the general dictionary corresponds to subentry 2, because both the precision of the extension of the semantic argument (‘players of a given sport’) and its intension (‘a given number of elements’) are higher. Furthermore, the group should have a given purpose that should give rise to a given value in the Telic Qualia structure (Pustejovsky, 1995).

The definition of the specialized value of talde ‘group’ as a mathematical term (Definition 1) obviously presents an even more complex semantic configuration. Limiting our interpretation to a very superficial analysis, we can establish that the first argument has a significantly more precise extension than in general meanings (giturak matematikoak ‘mathematical structure’). A similar conclusion applies to the second and third arguments that are the constituent elements of the ‘group’ (multzo bat ‘a set’ and barne-eragiketa bat ‘an internal operation’). Furthermore, both the ‘set’ argument and the ‘internal operation’ argument present several entailment relations that lead to their definitions in the field of mathematics. In addition, the ‘internal operation’ argument must fulfill three properties, which in turn lead to the entailment relations corresponding to their definitions. We can thus conclude that the terminological values of talde ‘group’ in mathematics present a much more complex semantic configuration than general values of this lexical item, just as predicted by Adelstein (2004, 2007).

3.2 Paraterms

Concerning the specialized values of talde ‘group’ that can be characterized as paraterms, at least two types of semantic configurations can be distinguished. In the first type of semantic configuration the second argument is saturated inside the compound noun with N1, which is a term in the specialized domain, so that the compound noun acquires a compositional specialized meaning determined by the combination of both elements (17a). In this type of paraterm there is an argument relation between the two elements of the compound noun. Following the classification by Bosque & Picallo (1996) of relational adjectives as thematic and classificatory, Zabala (1997) postulates that the grammar system of Basque does not allow thematic or argument adjectives. Taking into account Zabala’s view, (17b) should be ungrammatical, although such constructions are frequently found in corpora.

The examples in (18) illustrate the second type of semantic configuration. The second argument of the noun talde ‘group’ is saturated into the lexical item, giving rise to a shadow argument that is interpreted as ‘human beings that are part of a study population’. In these cases, the compound term also acquires a compositional meaning derived from the combination of both elements, but the relation between the two elements is classificatory, and adjectives are allowed (Zabala, 1997).

(18)

(a.) arrisku-talde Med.  
  risk-group  
  talde  
  ‘risk group’

(b.) talde eniko Soc. Sc.  
  group ethnic  
  ‘ethnic group’

In the fields of chemistry and biochemistry, the noun talde ‘group’ activates two different specialized values that can be classified in the second type of semantic configuration, that is, a semantic configuration with a shadow argument. In the former specialized value, talde is ‘a set of chemical elements that are part of the periodic table’, and in the second, talde is ‘an atom or group of atoms that are part of an organic molecule and are responsible for the chemical reactivity of such a molecule’. Therefore, the shadow arguments (sh-arg) of each value are different: sh-arg = chemical elements; sh-arg = atom(s). Moreover, the predications in which those arguments take part are also very different. In the first meaning they are very similar to those of general values: ‘to be constituted by several elements’ and ‘to share some properties’. This specialized value gives rise to some terms with a genitive (19a) or an ordinal (19b) modifier.

(19)

(a.) gas nobleen taldea  
  gas noble-GEN.pl group-DET  
  ‘the group of noble gases’

(b.) taula periodikoaren hamazortzigarren taldea  
  table periodic-GEN eighteenth group-DET  
  ‘the eighteenth group of the periodic table’

The semantic configuration of the second meaning is more distant from the general values of talde and it is more complex. First of all, the components of the group can be one or several atoms. Secondly, the talde ‘group’ must be a constitutive part of an organic molecule. And finally, it is the whole group (the first argument) that shares functional properties and not the elements of the group among each other. Terms bearing the element talde can focus on different elements of its semantic configuration. In (20a) the noun talde is combined in apposition with a noun that refers to a kind of functional atom group found in organic molecules (hidroxilo ‘hydroxy(l)’), establishing an hyponymy...
relation. A variant of the same term is obtained using the chemical formula of such a group (OH). The term in (20b) is a hyperonym of all the other terms in the paradigm, which focus on the characteristic that they share, that is, their chemical function. The term presents two variants: the variant calqued from other languages (talde funtzional [N + Adj.] lit. group functional ‘functional group’) and a more genuine variant (funtzio-talde [N+N] ‘function group’). The term in (20c) activates the semantic configuration value related to the function of the atom group (electron donor function). Finally, the term talde prostetiko ‘prosthetic group’ in (20d) presents an even more complex semantic configuration, as can be observed in Definition 3. This specialized value entails three features of the group of atoms: a) the type of organic molecule (enzyme) in which the group of atoms is a constituent, b) its structural characteristics (non-protein part = not constituted by amino acids), and c) its functional characteristics (the reactive part of the enzyme).

Definition 3: ZT Hitzea (Elhuyar 2009, 1268)

**talde prostetiko**: Biokim. Entzimen zati proteikoari gorog lotzen zaion molekula ex-proteikoa. Entzimen osagar iaunkorra da, eta ezinbestekoa da entzimek funtziona dezaten.

‘prosthetic group’: Biochem. Non-protein molecule that binds tightly to the protein part of enzymes. It is a constant part of enzymes, and it is necessary for the protein to achieve its function.’

In this paper we claim that the LNs in the examples (17–21) called paraterms can be considered Support Nouns following the terminology in Simone & Masini (2007, 2014). Nevertheless, Italian, French, Spanish, and English support nouns with general uses studied by Sinome & Massini are nouns with a high RF that decrease dramatically in certain combinatory contexts. In these contexts such nouns acquire an aspectual value that they project onto the noun with which they are combined: *atto di cortesia* (It.) lit. act of courtesy ‘kindness, courtesy’, *crisi de tosse* (It.) lit. crisis of cough ‘fit of coughing’, *coup de fil* (Fr.) lit. hit of wire ‘ring’, *golpe de suerte* (Sp.) ‘stroke of luck’, *fit of anger* (En.), *burst of enthusiasm* (En.)... Simone & Masini (2014) characterize these structures as in (I):

(I) Characterization of Support Noun structures (Simone & Masini 2014, 61)

Form: [N1 of N2] where N1 = support Noun

Semantics: ‘a single instantiation of an event that has to do with N2’

Furthermore, these authors locate Support Nouns at the highest level of the referentiality/nouniness scale, near the ultranouns. In this paper we argue that the aspectual values described by Simone & Masini are only some of the values that can present Support Nouns. In fact, in specialized usages of the language some nouns function as a support for the creation of complex terms. Such nouns are called paraterms (Estopà & Lorente, 2008, 2010).


However, some of the paraterms collected by Estopà & Lorente are clearly LNs, and their Basque equivalents behave similarly to the LNs analyzed in this paper: kantitate ‘quantity’, kopuru ‘number’, maila ‘level’, klase ‘class’, mota ‘type’, multzo ‘set’, talde ‘group’, and zati ‘part’.

Estopà & Lorente (2010) propose a functional definition of “paraterm” that is coherent with their applied aim to improve automatic extraction of terminology for their object language (Spanish).

(II) Definition of “paraterms” (Estopà & Lorente 2010, 302-303)

«Un paratérmino sería una unidad léxica, referida a conceptos básicos y con un alto grado de vaguedad semántica, usada en todos los ámbitos temáticos discursivos, incluído el general y todos los especializados, que combinado con un adjetivo con valor especializado específico puede constituir una unidad terminológica sintagmática en un determinado ámbito discursivo de especialidad.»

‘Paraterms are lexical items that refer to basic concepts and present a high level of semantic vagueness. They are used in every thematic discourse domain, including general and specialized discourses, and combined with an adjective bearing a specific specialized value, they give rise to phrasal term units in a particular specialized discourse domain.’

Our study on the activation of specialized values of LNs allows us to reformulate some elements in the definition (II) from both a theoretical viewpoint and taking into account the Basque language, which is typologically different from Spanish. First of all, we assume that a “paraterm” is not a lexical item, but rather one of the specialized values of a particular lexical item that can present other specialized and general values. Paraterms would be Support Nouns that contribute to the creation of complex terms (compound or phrasal terms), and that, in such combinatory contexts, would activate specialized values in their lexical configuration, depending both on the specialized theme in which they are used and on the lexical element with which they are combined. Therefore, we claim that paraterms are “specialized support nouns”. In this sense, they would differ from the Support Nouns described by Simone & Masini (2007, 2014), which habitually present aspectual values and should be considered as “general support nouns”. As for the referentiality/nouniness scale, general support nouns should be situated lower than specialized support nouns, since the values activated by the latter must be considered fully nominal. Consequently, the definition (II) should be reformulated as in (III):

(III) Reformulation of the definition of “paraterm” taking into account our analysis

A paraterm is one of the values in a lexical entry that belongs to the class of LNs, or at least presents a high level of semantic vagueness, and that activates particular specialized values into a specialized lexical item. This activation is due to certain factors in the communicative situation, clausal factors, and syntactic-semantic factors. Paraterms are “specialized support nouns” that, combined with a noun or with an adjective with specialized value, give rise to monolexical complex terms or phrasal terms.

Estopà & Lorente (2010) point out that among the terms made up of paraterms, one can distinguish “umbrella terms” (Estopà, 1999), which are hyperonyms of other specific terms, and also full terms. For instance, some examples used in this section, such as talde funtzional ‘functional group’ (20b), can be considered “umbrella terms”, whereas others, such as hidroxilo talde ‘hydroxy group’ (20a), can be considered specific terms. Going back to the referentiality scale in Simone & Masini (2010), specific terms would be located higher up that scale than umbrella terms.

Summarizing this section, we can conclude that the activation of specialized values in LNs always implies an increasing of the RF, which as predicted by Adelstein (2004, 2006), leads to higher referential specificity and greater imbrication of semantic information. Nevertheless, there would be different levels of strengthening referentiality, since terms are full lexical units whereas paraterms are support nouns that activate specialized values compositionally when combined with other lexical items to give rise to complex terms. The resulting terms can be umbrella terms, that is, hyperonyms of other more specific terms, or full specific terms, which for their part can be hyperonyms of certain umbrella terms. Consequently, umbrella terms are positioned lower down the referentiality scale than specific terms.

4 Conclusions

In this paper we have demonstrated that the hypotheses of Simone & Masini (2014) about the referentiality scale of nouns can be used to describe the behavior of Basque light nouns (LNs). Using a set of tests similar to those used by Simone & Masini, we have demonstrated that in Basque there are some types of nouns that can be considered LNs and that there are also referential force modulators (RFMs) which can increase or reduce the RF of nouns.

Secondly, we have shown that the terminologization of Basque LNs makes use of such RFM mechanisms in order to activate the specialized values required in different specialized communicative situations. As predicted by Adelstein (2004, 2007), specialized values of LNs present higher referential
specificity and greater imbrication of semantic information than general uses.

Finally, we have defined paraterms as specialized support nouns that are positioned at a lower level on the referentiality scale than terms, since in order to activate specialized values, paraterms are dependent on other lexical items, with which they combine to give rise to complex terms.

References


Notes

1. The work reported here was carried out within the project “EXTREMC: EXTracción de RELaciones entre Conceptos Médicos en fuentes de información heterogéneas” (TIN2013-46616-C2-1-R), which was supported by the MINECO (Ministry of Spanish Government).