



# Terminological Variation in Basque: Analysis of Texts of Different degrees of Specialization

Agurtzane Elordui, Igone Zabala

## ► To cite this version:

Agurtzane Elordui, Igone Zabala. Terminological Variation in Basque: Analysis of Texts of Different degrees of Specialization. *SKY Journal of Linguistics*, Linguistic Association of Finland, 2005, 18: 2005, pp.71-91. <artxibo-00000102>

**HAL Id: artxibo-00000102**

**<https://artxiker.ccsd.cnrs.fr/artxibo-00000102>**

Submitted on 11 Apr 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Agurtzane Elordui and Igone Zabala**

## **Terminological Variation in Basque: Analysis of Texts of Different Degrees of Specialization<sup>1</sup>**

### **Abstract**

The aim of this paper is to contribute to the evaluation of the present state of lexical-discursive development in scientific texts written in Basque. More specifically, it focuses on the study of terminological variation in texts on genetic engineering of different degrees of specialization. It describes the patterns of variation found in those texts, and attempt to relate such patterns to different types of sociolinguistic and discourse factors. The sociolinguistic factors are mostly associated with the incomplete normalization of Basque. For that reason, we also discuss the issue of terminological variation in languages undergoing normalization and we propose a set of recommendations based in this study, which we feel should be borne in mind during the normalization of terminology of those languages.

### **1. Introduction**

When a language has been undergoing a process of normalization for various decades, as in the case of Basque, it becomes necessary to evaluate how well the objectives set at different language planning stages have been met. There are numerous aspects that such an evaluation should cover, but our interest focuses specifically on analyzing the present state of lexical expansion and discursive development in fields where Basque was not used until recent times.

This general interest has led us to analyze the lexical-discursive development in science, as one of the areas of Basque language use to be incorporated most recently. Within this broad field, we have chosen to concentrate on the subject of genetic engineering due to its wide presence in texts of different degrees of specialization, ranging from newspaper and

---

<sup>1</sup> This work is part of the investigation project EHU 1/UPV 00113 310-H-15921/2004 developed in the University of the Basque Country.

magazine articles to high school and university textbooks or encyclopedia articles.<sup>2</sup> The functional variety of these texts enables us to analyze lexical and discursive development in different communication contexts, examining in particular whether or not the texts contain linguistic features that point to progress in the normalization of specialized discourse. Specifically, we have studied the degree of systematic variation developed to date in written Basque. Moreover, since the texts chosen for our study form part of the input required to build the science-related lexical skills and discourse strategies of different types of speakers—including secondary and university students, scientists and the general public—they enable us to analyze how these texts contribute to the lexical knowledge shared by the linguistic community, thus helping to make the community more compact.

The aim of this paper is to contribute to the evaluation of progress in scientific writing in Basque through a descriptive study of terminological variation. More specifically, it describes the patterns of variation found in different texts and attempts to relate such patterns to different types of sociolinguistic or discourse factors. To further the reader's comprehension of the sociolinguistic factors at work, the first part of this paper will be devoted to a brief description of the ongoing normalization process under way since the 1960s, when Basque began to be standardized. In describing the current situation, we will briefly explain the major sociolinguistic problems faced by the Basque linguistic community and how they affect the scientists and writers producing the texts concerned in this study. Sections 3 and 4 will then be devoted to terminological variation. First we discuss the issue of terminological variation in languages undergoing normalization, and argue that variation should actually be an objective to strive for within the normalization process. We then present the results of our descriptive study of terminological variation in Basque. We discuss each of the variation patterns observed in our corpus of Basque texts of different degrees of specialization. First we show the patterns of designative variation identified, and then attempt to account for each

---

<sup>2</sup> The texts examined are: texts from the encyclopedic dictionary *LUR Hiztegi Tematikoa* (~ 16,000 words); Basque high school texts from different publishers (~ 15,600 words); articles appearing in the science magazine *Elhuyar* (~ 4,000 words) and newspaper articles appearing in *Egunkaria* and *Berria* (~ 10,000 words). To help complete our study of the use of this terminology at higher levels of specialization, we consulted directly with experts in genetics at the University of the Basque Country, who kindly provided teaching materials used in fifth year genetics classes.

through explanations of the underlying sociolinguistic and discourse factors that may have motivated them. Finally, we propose a set of recommendations based on this study, which we feel should be borne in mind during the modernization (*cultivation or elaboration*) of languages undergoing normalization and, more specifically, in the normalization of terminology in Basque.

## 2. Language normalization and scientific writing in Basque

In the middle of the 20<sup>th</sup> century, Basque was in a state of diglossia and incipient displacement to Spanish and French.<sup>3</sup> Relegated for centuries to use almost entirely in the home,<sup>4</sup> it was practically absent from domains of public use, particularly areas concerned with technology and science. Written Basque was in practice almost nonexistent for common speakers, since only religious, and to a lesser extent literary, texts were available, and then only in very small numbers.

What is more, even this functional compartmentalization was breaking down to the detriment of Basque, with Spanish and French tending to be used more and more with family and friends. The lack of prestige and functionality of the language were, together with other political, cultural and economic factors (Elordui 1995), major reasons for abandoning use of the language. In a period of mass migration to cities and industry, with literacy on the rise and formal language domains gaining in importance — particularly those areas concerned with technology and science— Basque was not considered adequate for meeting the new functional requirements of Basque speakers (Elordui 1995). Moreover, the dialect fragmentation afflicting Basque through the centuries, together with the narrow range of text types, seriously hindered readership and the development of the language in formal written contexts.

In the 1970s, however, a new and promising period began for Basque, particularly as concerns the development and diffusion of the language in

---

<sup>3</sup> Basque is spoken in territories to the north and south of the Spanish-French border. In the middle of the 20<sup>th</sup> century, there was still a large proportion of monolingual Basque speakers on both sides of the border. Today however, in both cases the Basque-speaking community is mostly bilingual (Spanish-Basque/French-Basque) and shares the territory of the Basque Country in the south with a monolingual Spanish-speaking community, and in the north with a monolingual French-speaking community.

<sup>4</sup> Basque was also commonly used in social activities of a religious nature and at work among farmers and fishermen.

formal domains. The three most important factors accounting for this change were: the codification and standardization of Basque;<sup>5</sup> its introduction into the educational system,<sup>6</sup> in many cases as the vehicular language; and the use of Basque in the media. Importance must also be given to the progress and development that Basque has enjoyed in other public contexts (government, health, technology, etc.). According to a Basque Government sociolinguistic survey conducted in 2001, the more public services offered in Basque, the greater the demand for this language in such domains (*www.euskadi.net*).

These facts, together with other standardizing and normalization initiatives undertaken by individuals and institutions, have made it possible for written Basque to develop and for the language itself to spread to new areas of use. The result has been a spectacular increase in the publication of texts of different functional types and the creation of a growing community of consumers of formal written and oral texts in Basque.<sup>7</sup>

This situation, while undoubtedly promising, is somewhat offset by the difficulties encountered daily by Basque speakers in different usage areas. Moreover, apart from these problems common to all speakers of the language, the communities that produce texts in specialized fields suffer the particular difficulties associated with developing a type of discourse that was until lately non-existent in the linguistic repertoire of Basque. Finally,

---

<sup>5</sup> At the congress held in Arantzazu in 1968, the Royal Academy of Basque Language (*Euskaltzaindia*) embarked on a process of codifying and standardizing Basque in order to overcome the diglossia to which the language was subjected and curb its displacement (Lekuona 1968). The design of standard Basque was based primarily on sociolinguistic criteria (Mitxelena 1968) which sought not only to curb abandonment of the language and gain new speakers, but also to endow the language with linguistic resources enabling its use in new areas.

<sup>6</sup> In France, Basque does not enjoy recognition as an official language, and therefore its presence in education is limited to private schools. In Spain, Basque enjoys official status in the Basque Autonomous Community and in the northern part of the Autonomous Community of Navarra. In regions where Basque is an official language, schools offer students the possibility of studying in Basque (linguistic immersion model), or else studying in Spanish with Basque taught as a subject. The fastest-growing of the two is the immersion model, where Basque is the vehicular language for most subjects. The language is also gaining ground at the university level, although to a lesser extent. It is worth noting in the context of this article that a degree course taught entirely in Biology has been offered for more than a decade.

<sup>7</sup> The Basque community gained 110,000 speakers in the decade between 1991 and 2001. What is more, the population segment accounting for most of this increase is children and young people (*www.euskadi.net*), the prime consumers of formal writing.

we must not overlook the fact that the development of specialized language has occurred simultaneously with the ongoing codification and standardization of Basque, which has made it all the more difficult for specialized language to develop and become established.<sup>8</sup> Indeed, deficiencies in standardization affect these functional registers more than the registers reserved for interactions with family and friends. This is so because most specialized languages take the formal standard style as their reference point, which in the case of Basque is *euskara batua* or unified Basque.

One of the main obstacles to be overcome by people wishing to speak or write about science in Basque is the lack of a model to use in developing discourse. Until now, in most fields of study, Basque has had no texts that could be used as a reference and from which people could learn proper discourse strategies for that field. Basque scientists therefore have had no other option than to create their own scientific discourse, and in nearly all cases have had to deal with the problems arising from having, as their only reference, some other language, normally Spanish, French or English. To top things off, the option to borrow terms from one of these languages has been severely frowned upon, there being strong sociolinguistic pressure to create neologisms instead so as to shield the language from outside influence.

Another problem for the development and diffusion of scientific language is that the Basque science community is not a compact one. Communication networks among scientists are rare and discontinuous. Exchanges in Basque about science are mainly limited to academic contexts and then mostly to teacher-student relations.<sup>9</sup> This gives rise to two further problems affecting the development of scientific language. First, it seriously hinders the development and use of scientific Basque at higher levels of expertise, since it is through interactions among experts that specialized terminology is created. Secondly, the fact that specialists fail to share terminology and discourse strategies in their field means that the input received by their students —the potential future members of the science community— will, instead of making the community more

---

<sup>8</sup> Specifically, the first volume of the Unified Dictionary of the Basque Language was published in the year 2000 (Euskaltzaindia 2000).

<sup>9</sup> In these spheres as well, the input received by students is produced by teachers and professors whose own access to specialized knowledge has been mostly through Spanish or English. It would be interesting to study how this fact has affected the development of specialized discourse to date.

compact, foster further free variation in the community's use of terminology.

Basque, therefore, does not yet have a compact community of scientists sharing a common fund of scientific discourse to which they can turn for guidelines when producing new text. That is, there are no discourse strategies or established terminologies shared by this speech community. The creators of the specialized texts in our corpus find themselves without reference models, as happens as well with authors of teaching materials and articles intended for the general public. In the case of textbooks and teaching materials, there is the additional problem that a large number of them are translations from Spanish.

Having explained some of the major problems affecting producers of specialized texts and in general those involved in modernizing Basque, brief mention should be made of the important contribution that the type of texts chosen for this study can make in the development of specialized vocabulary or terminology in Basque. As pointed out above, the genetic engineering texts analyzed are intended for different levels of expertise and help in different ways to make the language community in this field more compact. Textbooks and teaching materials provide students with the terminology that will eventually enable them to share their specialized knowledge, and that will serve as the first rung on the ladder leading to higher and higher levels of specialization. The terminology acquired during these early stages is therefore essential to the building of terminology in any field of knowledge. The lack of a shared standard lexicon at these basic levels makes it all the more difficult for normalized terminology to exist at higher levels of specialization. Consequently, these texts are crucial for the contribution they make to the degree of compactness of the discourse community in each field of specialization. Encyclopedic articles, on the other hand, help in that they both reflect and secure the encyclopedic-lexical knowledge of a linguistic community. They can play an important role during phases in which the language is not sufficiently normalized or stabilized in certain areas of study, establishing terminology and offering it besides in a more or less educational context. Finally, texts intended for the general public extend some of the vocabulary of different specialties among the linguistic community. Such texts are particularly important for filling in the gaps for speakers of minority languages excluded from the educational sphere, and they therefore can also play a large role in making the speech community more compact.

### 3. Terminological variation in normalized languages

Terminology can be approached from a standardizing, normalizing or descriptive standpoint, and each approach sees variation in a different light. Followers of the Wüster school (Wüster 1998 [1979]) and the Soviet school (Lotte 2001 [1948]) attempt to combat synonymy through standardization, since they feel that the proliferation of variants hinders communication among specialists. The normalization school centered around the recuperation of French in Quebec sees synonymy as a linguistic variation phenomenon that must be addressed, but not necessarily eradicated (Auger 1994, Boulanger 1983). Finally, Socioterminology and Communicative Theory — the descriptive schools of terminology — criticize the Wüster school (General Theory of Terminology, GTT) for its reductionism primarily as far as the tendency to obviate variation in terminological units is concerned. Socioterminological (Boulanger 1991; Gaudin 1993; Guespin 1991) and communicative (Cabré 1999) approaches, on the other hand, see terminological variation as one of their main subjects of study. They describe terminological units in discourse and analyze the sociological and discourse conditions that give rise to different types of texts. These approaches enable us to relate inter- and intratextual variations in terminology to the degree to which discourse has developed in the language.

In line with these latter schools, we assume that there is designative and conceptual variation in terminological units within and between texts, and that these two types of variation depend both quantitatively and qualitatively on the characteristics of the text, particularly on its degree of specialization (Freixa 2001, 2003). In a given specialized discourse community, there is implicit or explicit agreement on the notion-designation relationship established within each terminological unit, although the discourse community also shares a system of variation that allows a) the terminology to adapt to communicative situations of different degrees of specialization and abstraction, and b) specialized discourse to be ordered using different discourse strategies (paraphrase, reformulations, metaphors, etc).

In analyzing variations in different types of texts, we must first draw a distinction between variants found in texts written by the same author (individual or intra-variation) and variation in texts written by different authors (inter-variation) (Freixa 2003). In the case of individual variation, stylistic factors (e.g. the desire to avoid repetition, or find more expressive,



succinct, emphatic or erudite ways of expressing something) often underlie designative variants and should therefore be taken into special account. Individual speakers may also vary the expressions they use, including scientific terminology, for functional reasons (e.g. to adapt to their listener's level of specialization or the degree of formality required by the occasion). All such causes are discursive in nature. Generally speaking, efforts to enhance precision, systematicity and conciseness can be seen in the linguistic choices made by the authors of texts of different degrees of specialization.

However, in comparing texts written by different authors, other types of causes may be at work. Variations can occur due to dialect differences of geographical, chronological or social origin, or they may arise for cognitive reasons associated with e.g. concurrent theories, conflicting concepts of different fields of study, etc. Geographical dialect differences are found primarily in texts produced in different countries, whereas chronological dialect variation mostly affects developing lexicons where the level of established vocabulary is still very low (Guilbert 1973).

#### **4. Patterns of terminological variation in Basque**

According to a number of studies, the patterns of variation found in sociolinguistically unstable, minority languages that have not been fully normalized such as Basque, differ considerably from those found in normalized ones. The two major distinctions are a greater frequency of free variation and the difference in factors accounting for this variation (Dorian 1981; Dressler 1981; Elordui 2003; King 1989). With these findings in mind and on the premise that the more than 25 years of use of scientific language in academic settings and the press must have produced some progress in discourse, in our study we formulated the following hypotheses: a) the terminological units of the corpus will exhibit variation patterns which can, as in normalized languages, be accounted for by discourse factors, b) but there will also be other variation patterns that can only be accounted for by sociolinguistic factors stemming from the current instability of Basque. This latter type of variation will be highly unsystematic.

In this article we are concerned with designative variation — i.e., the phenomenon of designating in different ways a single item having specialized meaning. Designative variation is much easier to detect than conceptual variation, and besides is often reflected directly in the

alternative designation itself. Classified under this heading are lexical variants (variants that could be considered synonyms), and graphic, spelling and morphosyntactic variants.

Designative variants were extracted manually from the corpus and classified according to morphosyntactic pattern.<sup>10</sup> The resulting variation patterns were analyzed from two standpoints: a) intratextually, comparing variants within a single text to analyze the author's fund of terminology and discourse strategies, and b) intertextually, comparing variants found in different texts to find patterns of variation motivated by different discourse conditions (correlating variants to degree of specialization, author's objectives, etc).

#### 4.1 Analysis of designative variation patterns

To identify and analyze morphosyntactic variation patterns, we used the classification proposed by Freixa (2001, 2003), adjusted to suit the characteristics of Basque and supplemented by other patterns not included in Freixa's classification. Examples of the patterns found are set out below:

- a) Graphic variants: spelling variants (*ingeniaritza genetiko/injinerutza genetiko* 'genetic engineering'); variant order of letters in abbreviations (*ADN/DNA*); hyphen-space alternation (*DNA-kate/DNA kate* 'DNA chain'); typographical alternation of quotation marks («*ADN zunda*»/*ADN zunda* 'DNA probe'); term-acronym alternation (*Polimerasaren katea-erreakzioa/PCR* 'polymerase chain reaction'); term-abbreviation alternation (*ARNm/ARN mezulari* 'messenger RNA'); capital-small letter alternation (*Giza Genoma Proiektua/Giza genoma proiektua* 'Human Genome Project').
- b) Morphosyntactic variants: subordinant compound N1-N2/attributive compound N2 N1 alternation (*ama-zelula/zelula ama* lit. mother-cell/cell-mother 'stem cell'); N-N compound/PP+N sequence alternation (*zelula-nukleo/zelularen nukleoa* lit. cell-nucleus/cell-

---

<sup>10</sup> Automatic extraction tools for Basque are still in their infancy and took their first steps only recently in the field of variation (Alegría *et al.* 2004), focusing almost exclusively on morphosyntactic patterns. The relative scarcity of technical dictionaries and texts hinders statistical methods, and for the moment lack of knowledge of discourse markers makes it impossible for automatic detection tools to apply them.

GEN nucleus ‘cell nucleus’); N-N compound/N+Rel.Adj. type syntactic sequence alternation (*gene-terapia/terapia geniko* lit. gene-therapy/therapy genic ‘gene therapy’); N-N compound / derivative alternation (*landare-hobekuntza/fitohobekuntza* ‘plant improvement/phytoimprovement’); N+Adj.<sup>11</sup>/PP+N alternation (*gaixotasun hereditario/herentziazko gaixotasun* lit. disease hereditary/heredity-GEN disease ‘hereditary disease’).

- c) Reduction: of extension (*izaki bizidunak/izakiak* lit. organism life-having/organisms = ‘(living) organisms’); of base (*izaki bizidunak/bizidunak* lit. organism life-having/life-having ‘(living) organisms’); other types of reduction changing the category of one of the elements (*eragile kartzinogeniko/kartzinogeno* lit. agent carcinogenic/carcinogen ‘carcinogenic agent/carcinogen’); or Basque term vs. borrowed term alternation (*zelula hartzaile/errezptore* lit. cell receptor/receptor ‘receptor cell/receptor’).
- d) Lexical variants: between two simple Basque terms (*txertatu/erantsi* ‘insert’) or a simple vs. a complex Basque term (*hari/harizpi* lit. thread/ thread-strand ‘strand’); borrowed term vs. complex Basque term (*bektore/garraiatzaile* ‘vector/transporter’; *organismo/bizidun* lit. organism/ live-being ‘organism’); scientific term vs. Basque vernacular (*zigoto/obulu ernaldu* ‘zygote/fertilized egg’); in components of complex units, such as variation in prefix (*birkonbinatu/berkonbinatu* ‘recombine’) or suffix (*klonatze/klonaketa* ‘cloning’), variation in nucleus of a compound (*DNA zati/DNA segmentu* lit. DNA piece/ DNA segment ‘DNA segment’), variation in compound modifier (*hazkuntza-ingurune/kultibo-ingurune* ‘growth medium/ culture medium’); in polylexical units such as variation in the base (*aztarna genetiko/arrasto genetiko* lit. footprint genetic ‘genetic footprint’), variation of extension (*aztarna genetiko/aztarna geniko* lit. footprint genetic/ footprint genic ‘genetic footprint’); *gene azpirakor/gene errezesibo* ‘recessive gene’; *transkriptasa alderantzikatua/alderantzizko transkriptasa* lit. transcriptase reversed/reverse-GEN transcriptase ‘reverse transcriptase’); variation in base and extension (*bakterio-genoma/bakterioaren DNA* lit. bacteria-genome/bacteria-GEN DNA

---

<sup>11</sup> Note that in Basque, like in Spanish, adjectives are commonly placed to the right of the noun.

‘bacterial genome/ bacterial DNA’; *ernaldutako obozitoa/obulu ernaldua* lit. fertilized-GEN oocyte/oocyte fertilized ‘fertilized oocyte’).

As noted in previous sections, there are different motivating factors behind these variation patterns. Intratextual and intertextual analysis reveals some that appear to be due to discourse factors. However, many of the patterns detected in the two analyses cannot easily be accounted for by this type of functional motivation, and instead must be explained in terms of sociolinguistic instability. Examples of each type are discussed below.

#### 4.2 Variation patterns due to discourse factors

Intratextual analysis reveals that certain variation patterns are due to discourse factors. For example, authors often resort to graphic alternation to achieve conciseness, as is the case when they use both a term and its abbreviation in the same text: (*Polimerasaren katea-erreakzioa* ‘polymerase chain reaction’/*PCR*; *Giza Genoma Proiektua* ‘Human Genome Project’/*HGP*). Once an author has given the meaning of the abbreviated form, he uses it thereafter in the text for reasons of economy.<sup>12</sup>

Morphosyntactic variants can be explained in similar terms: e.g. compound/derivative alternation (*landare-hobekuntza* ‘plant improvement’/*fitohobekunta* ‘phytoimprovement’). The effort to be concise and avoid repetition leads authors to use forms of different degrees of succinctness, once they have established the meaning equivalence between variants. Different types of reductions (*izaki bizidunak* lit. organism live-having ‘live beings’/*izakiak* lit. beings ‘organisms’/*bizidunak* lit. live-having ‘organisms’) can be attributed to the same discourse and stylistic causes — i.e. the desire to be concise and avoid repetition.

Somewhat different are patterns involving lexical variants of different types: e.g. one Basque term vs. another (*txertatu/erantsi* ‘to insert’); simple vs. complex Basque terms (*hari* ‘thread’/*harizpi* ‘thread-strand = strand’); Basque term vs. borrowed term (*garraiatzaile* ‘transporter’/*bektore* ‘vector’); or Basque vernacular vs. scientific term (*obulu ernaldu* lit. ovule fertilized ‘fertilized ovule’/*zigoto* ‘zygote’). Underlying such instances is

---

<sup>12</sup> On some occasions authors resort to a reference language other than Basque for abbreviations (*PCR*), while on other occasions, the abbreviation stems directly from the Basque sequence (*GGP*).

also the desire to avoid repetition, to sound more erudite or to improve style. Borrowed and scientific terms vs. their vernacular counterparts, and complex vs. simple forms<sup>13</sup> are all accorded greater prestige within the linguistic community.

There is a different type of discourse reason for quotation marks variation («*ADN zunda*»/*ADN zunda* ‘DNA probe’). This pattern seems to be due to the author’s desire to establish the terminological value of a given sequence within the text: the first time the term appears, the author sets it off in quotation marks, but does not use them thereafter once the terminological value of the sequence has been established.

Finally, morphosyntactic alternation such as ‘compound vs. syntactic sequence’ or alternation between different types of syntactic sequence may often be due to syntactic-semantic requirements of the language itself. For example, alternation between a compound and a syntactic sequence of the type PP+N (*zelula-nukleo* ‘cell nucleus’/*zelularen nukleoa* cell-GEN nucleus ‘cell’s nucleus’) is essential for distinguishing between generic and specific reference. In Basque, the modifier in subordinative noun compounds always has generic reference, unless specific reference is shown syntactically by adjoining a case marker and article, giving rise to a PP+N noun phrase. Another example would be PP+N vs. N+Rel.Adj. (*herentziazko gaixotasun* lit. inheritance-GEN disease ‘hereditary disease’/*gaixotasun hereditario* lit. disease hereditary ‘hereditary disease’), an alternation that is often due to the requirement that coordinated elements must belong to the same syntactic category: *gaixotasun hereditario eta sexualak* lit. disease hereditary and sexual ‘hereditary and sexual diseases’/*herentziazko eta transmissioko gaixotasunak* lit. inheritance-GEN and transmission-GEN diseases ‘hereditary and transmitted diseases’.

Analysis from the intertextual point of view reveals a rise in established discourse rules associated with a text’s degree of specialization and the discourse community to which the author belongs. For example, the presence or absence of hyphens in dependent noun compounds can show whether or not the author belongs to the science community. The standard code allows subordinative noun compounds to be written with or without a hyphen (*telefono-zenbaki/telefono zenbaki* = ‘telephone(-)number’) and this

---

<sup>13</sup> The choice of a Basque complex form rather than a simple form as a technical term is particularly evident in the case of tautological compounds where both elements have the same semantic content (*larruazal* ‘skin-skin’, *lurzoru* ‘ground-ground’), but it is frequent as well in other compounds such as *gainazal* ‘above-surface = surface’.

has given rise to different stylistic options. Novelists and journalists tend to avoid this use of the hyphen as much as possible, whereas scientists feel that it not only can, but should be used for the sake of accuracy. So, for example, they write *gene-emaile* ‘gene-donor = donor of genes’ with a hyphen to show that they are referring to a type of donor, whereas they write *zelula hartzaile* (lit. cell receptor = ‘receptor cell’) without a hyphen to show that they are referring to a type of cell — i.e. one that is a receptor (of genes). The textbooks and *Elhuyar* magazine articles analyzed for our study were found to follow the scientists’ policy. The encyclopedic dictionary *LUR*, on the other hand, and newspaper articles normally do not use the hyphen. We conclude therefore that the presence of hyphens in this type of compound indicates the discourse community to which the author or editor of the article belongs.

When we compared different types of texts written by authors all clearly belonging to the same discourse community, since they systematically used the hyphen in subordinative noun compounds, we found inconsistencies in its use with certain other types of compounds (*DNA sekuentzia/DNA-sekuentzia* ‘DNA(-)sequence’; *DNA kate/DNA-kate* ‘DNA(-)chain’). Interviews with these authors revealed that such differences are due to conceptual variation. That is, those that use the hyphen see a dependency relationship in the compound (‘the sequence/ the chain is formed by DNA’), while those who leave it out see a relationship of apposition (‘the sequence/ the chain is (called) DNA’).

### 4.3 Variation patterns due to sociolinguistic factors

Many of the terminological variation patterns found in the corpus were not motivated by functional causes or the exigencies of discourse, but arose rather from factors of a sociolinguistic nature. We feel that the consolidation of specialized terminology is influenced mainly by four such factors. a) First, the instability of the standard code, plus b) the linguistic dependency arising from the language’s minority status, and its hitherto infrequent use in specialized contexts. c) This latter factor causes variation particularly if combined, as in the case of Basque, with strong sociolinguistic pressure on the part of normalizing agents. d) Finally, to the free variation arising from all these factors must be added the lack of fluid communication networks which could contribute to the fixing of specialized vocabulary.

a) Some types of graphic variation are associated with the instability of the code itself. The spelling differences we detected are not due to geographical dialect, as happens in languages such as Spanish or English, which are spoken in vast, well-differentiated geographical areas. In the case of Basque, these variation patterns include recently standardized variants (*kate* ‘chain’, *ingeniaritza* ‘engineering’, *zientzialari* ‘scientist’, *xurgatu* ‘to absorb’) vs. their non-standard variants (*katea*, *injinerutza*, *zientzilari*, *zurgatu*), which were used extensively in the linguistic community prior to the recommendations of the official unified dictionary *Hiztegi Batua* (Euskaltzaindia 2000). Our analysis of the corpus shows that the non-standard variants appear particularly in textbooks and encyclopedia articles (all published before 2000), but are much less frequent in magazine and newspaper articles. Since the latter are published at regular intervals, they can adapt their style-sheets and update them more often, whereas textbooks and encyclopedias are re-edited only after a certain number of years. Therefore they remain in circulation long after their spelling and style-sheets have been rendered outdated by the publication of new standard recommendations.

b) As for linguistic dependency, our study of the corpus revealed a tendency to borrow lexical units from the reference language (mainly Spanish) and to retain their original structure. This gives rise to borrowed terms (*bektore*, *organismo*) and sequences patterned exactly after the Spanish (*terapia geniko* ‘terapia génica’ = ‘gene therapy’, *aztarna genetiko* ‘huella genética’ = ‘genetic footprint’, *plasmido birkonbinatzaile* ‘plásmido recombinante’ = ‘recombinant plasmid’).

A second consequence of linguistic dependency, and more concretely of the failure to use Basque in specialized spheres, is the clear lack of writing skills or discourse competence evident in some of the texts studied. Let us take for example cases in which it is necessary to construct or deconstruct complex terminological units by binding or separating the elements of which they are composed, depending on the exigencies of discourse. In Basque, speakers must be able, for example, to dissociate the items in a subordinative noun compound when one of the elements is modified. Basque nominal compounds are usually formed by two elements. The introduction of a third element asks for a postpositional nexus: *transmisio-gaixotasunak* lit. transmission-diseases ‘transmitted diseases’/ *transmisio sexualeko gaixotasunak* lit. transmission sexual-GEN diseases ‘sexually-transmitted diseases’. However, in our corpus we detected cases where authors failed to make the required change and simply formed a

sequence of three elements without any kind of postpositional nexus, thereby revealing a gap in discourse competence: *azido nukleiko harizpi* lit. acid nucleic strand ‘nucleic acid strand’; *gene-transferentzi sistema* ‘gene-transfer system’.

Problems with word formation also appeared to stem from poor discourse competence. Here is a small example:

- (1) *Proteina onkogenikoek minbizia sortzen dute proteina normalen funtzioa partzialki edo modu desegokian imitatzen dutenean, zelula normal bat zelula kantzeroso bihurtuz...*  
 ‘Oncogene proteins produce cancer when they partially or inappropriately imitate the function of normal proteins, thus turning the normal cell into a cancer cell ...’

In the explanation the author uses the well fixed Basque term. (*minbizi* ‘cancer’), instead of the loan word (#*kantzer*). However, when it comes to using the polylexical term ‘cancer cell’, the author resorts to the borrowed Spanish term (*zelula kantzeroso*), instead of following normal word-formation rules in Basque to produce *zelula minbizidun*. Such a solution would also have been much more consistent with the development of the discourse.

Deficient mastery of the discourse strategies available in Basque to achieve conciseness and avoid repetition is also evident from the way authors use reductions of the base or the extension in polylexical units, as shown in the following example.

- (2) *Bioteknologiak Injinerutza Genetikoaren eta beste alor batzuen (...) teknikak bildu eta izaki bizidunei ezartzen dizkie gizadiarentzat onurak lortzeko helburuz(...) Azkeneko bi kasuetan, organismo batean material genetikoa sartzeak haren eduki genetikoa aldatzen du.*  
 ‘Biotechnology combines the techniques of Genetic Engineering and other fields (...) and applies them to living beings to benefit human life (...) In the last two cases, the introduction of genetic material in an organism alters its genetic content.’

The author resorts to the borrowed term *organismo*, despite the availability of two reduced Basque forms (*izaki* ‘being = organism’ and *bizidun* (lit. life-having = ‘organism’) which would solve the author’s discourse objectives of conciseness and avoidance of repetition of the term *izaki bizidun* ‘living being’. Note that in Spanish there is no reduced variant for the polylexical term *ser vivo* (‘living being’) — i.e. \**ser* or \**vivo* — and that the only variant available is the monolexical term *organismo*.



Presumably the author of (2) simply copied this discourse strategy from Spanish.

c) As in other languages, and more particularly in minority languages where there is strong sociolinguistic pressure, there are, in addition to variants stemming from the reference language as described above, variants that have been proposed by the different normalizing agents for the purposes of counteracting lexical dependency on Spanish. A case in point is that of N + Rel.Adj. sequences,<sup>14</sup> which have been officially labeled foreign to Basque by the Royal Academy of the Basque Language (Euskaltzaindia 1992). From that moment on, numerous linguists, translators and editors have almost systematically replaced this type of sequence with noun compounds. Examples from our corpus include: *terapia geniko/gene-terapia* ‘gene therapy’; *plasmido bakteriano/bakterio-plasmido* ‘bacterial plasmid’; *ugalketa sexual/sexu-ugalketa* ‘sexual reproduction’.

Other cases of sociolinguistic pressure include the effort to avoid ill-formed derivatives in Basque through the formal recommendation of supposedly more appropriate terms. Here, as in the previous case, the consequence is terminological dispersion, occasioned by the coexistence of variants that emerged during different phases of linguistic intervention (Gutiérrez 1998: 98). As an example from our corpus, let us take the term *recombinant plasmid*, for which the following Basque variants exist: *plasmido birkonbinatzaile/plasmido birkonbinagai/plasmido birkonbinatu*.

d) As noted above, in addition to all these factors motivating variation, in the case of Basque there is also the further problem of a lack of fluid communication networks which could help to fix and consolidate specialized vocabulary. In normalized languages word formation processes (derivation and composition) tend to block when a lexical unit for a given notion already exists. In Basque, however, there is an overproduction of these formation processes. In our corpus we have encountered numerous examples of variation associated with this phenomenon: e.g.: *klonatze/klonazio/klonaketa* for the term *cloning*, or *ama-zelula/zelula ama* for *stem cell*.

In some cases, the Unified Dictionary makes a distinction between variants, as for example *haziera* ‘growth’/*hazkuntza* ‘offspring, culture’. However, the actual uses that we find in the texts in our corpus show that

---

<sup>14</sup> See Zabala (1997) on Basque equivalents of different types of Spanish relational adjectives.

the unification recommendations are not always known or followed by writers, who use such variants in free variation.

Finally, brief mention should be made of the fact that some authors of texts in our corpus are aware of the lack of an established, specialized vocabulary and attempt to offset this deficiency by resorting to certain discourse strategies. The following example illustrates one available resource — the use of quotation marks for metalinguistic purposes, in this case to show the reader that the term in quotes, *eraginpean* ‘effect-under = undergo’, is a fixed technical term and is not used by accident or for stylistic reasons:

- (3) *Sagu-kume hau terapia geniko baten «eraginpean» dago.*  
 mouse-baby this therapy genetic DET-GEN «effect-under» is  
 ‘This baby mouse is undergoing gene therapy.’

## 5. Conclusions

Analysis of the designative variation patterns found in our corpus confirms our opening hypotheses. In both intratextual and intertextual analyses we have found variation patterns motivated by discourse factors — i.e. variants within a single text due to different discourse requirements, and differences within a range of texts that are due to the variety of discourse conditions in each type of writing. These patterns show the progressive establishment of discourse rules both at the intratextual and intertextual levels.

However, we have also observed variation patterns that cannot be accounted for by discourse motivations and which seem to be due instead to sociolinguistic factors associated with the incomplete normalization of Basque. As was to be expected, such variants are mostly in free variation. However, we should note that we have also found systematic variation associated with incomplete normalization, more specifically, variation patterns showing that authors attempt to compensate in discourse for the lack of established terminology.

There are various sociolinguistic factors that can account for the free variation found in the corpus.

a) First, free variation is in many cases motivated by the relative instability of the standard code. As noted above, the use of Basque in specialized areas of study began years before the lexical standardization recommendations issued in the year 2000, a fact that has given rise to spelling and lexical variants fitting the ‘standard vs. nonstandard variant’

pattern. The texts studied show that the standard code made its way more quickly into newspaper and magazine articles thanks to their greater frequency of publication, whereas textbooks, which have a greater influence on the sector of the speech community most likely to enter fields of specialized study, have taken much longer to adapt to the code.

b) Other patterns of free variation appear to be due to the pull of opposing forces. On the one hand there is the linguistic dependency that comes from reference terms established in another language, while on the other there is the strong sociolinguistic pressure so often present in languages undergoing normalization, which prefers the creation of neologisms to borrowed terms or words patterned on the formation rules of other languages. The natural consequence of such a conflict is terminological dispersion. Different experts, translators and institutions come up with different solutions to terminology problems. Moreover, the lack of fluid communication networks between speakers in the same field of study is a further impediment. The consequence observed in our corpus is an overproduction of variants occurring in free variation.

c) Finally, free variation is also related to authors' linguistic proficiency and writing skills, and to their knowledge of the specialized vocabulary of the field in question. The corpus reveals a broad range of skill in these areas. Some authors tend to use variants stemming from the reference language and, as intratextual analysis shows, variants that reveal a certain lack of discourse competence. Such failings are particularly evident when the author attempts, for discourse reasons, to generate in Basque terms drawn from other languages that require the adaptation of components of polylexical items.

To conclude, this preliminary study of terminological variation in Basque has led us to a number of criteria that should be borne in mind when undertaking the normalization of a language such as Basque. Indeed, the first two take the form of recommendations specifically for this language, while the rest should, we feel, be borne in mind in any process of linguistic normalization:

1. Variation due to instability of the standard code appears likely to lessen as the code becomes stabilized. However, we feel that a normalizing initiative designed to update all the publications in the market, particularly textbooks and teaching materials, would help to provide the community with an input that would at least not foster free variation.

2. As for terminological dispersion, we assume that the normalizing activity of the Terminology Committee recently created by the Basque

Government Department for Linguistic Policy will help to establish reference terms, thereby attenuating the situation.

For any language currently in the process of modernization and seeking to normalize the terminology of specialized fields of study, we would make the following general recommendations: First, it is crucial to recognize that incipient systematic variation due to discourse reasons should be seen as grounds for promoting such variants, particularly in the academic world. Second, we would recommend pursuing a linguistic policy flexible enough to allow different users to contribute to the discursive development of the language. The results of our study show that in cases where the standard model is flexible (use of a hyphen in subordinative noun compounds), usage rules have emerged in different Basque discourse communities, whereas overly rigid normalizing recommendations (e.g. N+Rel.Adj. sequences considered foreign to the language) simply foster terminological dispersion. Finally, we consider that normalizing activity should focus not on the eradication of free variation, but rather on channeling it. Today's free variation could provide tomorrow's established repertoire of functional variants for different discourse requirements.

## References

- Alegría, Iñaki *et al.* (2004) Linguistic and Statistical Approaches to Basque Term Extraction. In *Actes de GLAT-Barcelona 2004*, pp. 235-246. Barcelona: Universitat Politècnica de Catalunya.
- Auger, Pierre (1994) Pour un modèle variationniste de l'implantation terminologique dans les entreprises au Québec. In *Colloque sur la problématique de l'aménagement linguistique: Actes du Colloque*, pp. 483-493. Quebec: OLF.
- Boulanger, Jean-Claude (1983) Synonymie, néonymie et normalisation en terminologie. In Duquet-Picard, D. (ed.) *Problèmes de la définition et de la synonymie en terminologie. Actes du colloque international de terminologie*, pp. 311-327. Quebec: Girsterm.
- (1991) Une lecture socioculturelle de la terminologie. *Cahiers de Linguistique Sociale* 18: 13-30.
- Cabré, Maria Teresa (1999) Elementos para una teoría de la terminología: hacia un paradigma alternativo. In Maria Teresa Cabré (1999) *La terminología. Representación y comunicación*, pp. 69-92. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Ciapuscio, Guiomar (2003) *Textos especializados y terminología*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Dorian, Nancy (1981) *Language Death: The Life Cycle of a Scottish Gaelic Dialect*. Philadelphia: University of Pennsylvania Press.

- Dressler, Wolfgang Ulrich (1981) Language shift and language death: a protean challenge for the linguist. *Folia Linguistica* XV/1-2: 5-27.
- Elordui, Agurtzane (1995) Hizkuntz heriotza eta aldaketa morfologikoa: Hegomendebaldeko bizkaieraren kasua. Ph Dissertation. University of the Basque Country.
- (2003) Variation in the Grammar of Endangered Languages: The Case of Two Basque Dialects. *SKY Journal of Linguistics* 16: 7-26.
- Euskaltzaindia (1992) *Hitz elkartuen osaera eta idazkera*. Bilbao: Euskaltzaindia.
- (2000) Hiztegi Batua. *Euskera* 2000, II volume. Bilbao: Euskaltzaindia.
- Freixa, Judit (2001) Reconocimiento de unidades denominativas: incidencia de la variación en el reconocimiento de las unidades terminológicas. In Maria Teresa Cabré & Judit Feliu (eds.) *La terminología científico-técnica: reconocimiento, análisis y extracción de información formal y semántica*, pp. 57-65. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- (2003) La variació terminològica. Anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient Ph Dissertation. Universitat Pompeu Fabra. Barcelona.
- Gaudin, François (1993) Pour une socioterminologie: des problèmes pratiques aux pratiques institutionnelles. Rouen: Publications de l'Université de Rouen.
- Guespin, Louis (1991) La circulation terminologique et les rapports entre science, technique et production. *Cahiers de Linguistique Sociale* 18: 59-79.
- Guilbert, Louis (1973) La spécificité du terme scientifique et technique. *Langue Française* 17: 5-17.
- Gutiérrez, Bertha M. (1998) La ciencia empieza en la palabra. Análisis e historia del lenguaje científico. Barcelona: Editorial Península.
- King, Ruth (1989) On the social meaning of linguistic variability in language death situations: Variation in Newfoundland French. In Nancy Dorian (ed.). *Investigating obsolescence. Studies in language contraction and death. Studies in the Social and Cultural Foundations of Language* 7, pp. 139-148. Cambridge: Cambridge University Press.
- Lekuona, Manuel (1968) Euskeraren batasuna. *The Conference of Arantzazu. Euskera XIII*: 139-141.
- Lotte, Demitrii Semenovich (2001[1948]) Condiciones para la precisión y la claridad de la terminología. In Maria Teresa Cabré, Judit Freixa, Mercè Llorente y Carle Tebé (eds.) *Textos de Terminología de la Escuela Rusa*, pp. 131-150. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Mitxelena, Koldo (1968) Ortografía. *The Conference of Arantzazu. Euskera XIII*: 202-219.
- Wichter, Sigurd (1994) *Experten-und Laienwortschätze. Umriss einer Lexikologie der Vertikalität*. Tübingen: Niemeyer.
- Wüster, Eugen (1998[1979]) *Introducción a la teoría general de la terminología y a la lexicografía terminológica*. Maria Teresa Cabré (ed.). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Zabala, Igone (1997) Argumentu-harremana/ eremu-harremana: izenondo erreferentzialen euskal ordainen bila. In *International Congress on Terminology*, pp. 535-556. San Sebastián: IVAP-UZEI.

Contact information:

Agurtzane Elordui  
Euskal Filologia Saila  
Gizarte eta Komunikazio Zientzien Fakultatea  
Euskal Herriko Unibertsitatea  
Sarriena auzoa z./g. 48940 Leioa  
Spain  
e-mail: agurtzane.elordui@ehu.es

Igone Zabala  
Euskal Filologia Saila  
Zientzia eta Teknologia Fakultatea  
Euskal Herriko Unibertsitatea  
Sarriena auzoa z./g. 48940 Leioa  
Spain  
e-mail: igone.zabala@ehu.es

