

INTERDISCIPLINARY TRAINING ASSESSMENT OF COMMUNICATION SKILLS FOR STUDENTS WITH BASQUE AS INSTRUCTION LANGUAGE IN THE FACULTY OF SCIENCE AND TECHNOLOGY AT UPV/EHU UNIVERSITY

I. Zabala, J.R. Aiartza, A. Apraiz, A. Aranburu, J.M. Arizmendi, N. Arrizabalaga,
A. Eiguren, J. Garcia, I. Martinez, J.C. Odriozola, M. Olazar, M.J. Omaetxebarria,
M. Ortiz, O. Oruetxebarria, N. Zabala, O. Zuloaga

University of the Basque Country (UPV/EHU) (SPAIN)

Abstract

One of the main goals of university education is to provide the students with the specialized communication patterns shared by experts in a disciplinary field, i.e. specialized text genres and linguistic registers characteristic of these text types. The end-of-degree projects are one of the main tools to assess the acquisition of communication skills by students. This paper describes the process of elaboration of an assessment rubric aimed to be used as a reference for students and instructors involved in the elaboration of end-of-degree projects in Basque. The rubric is based on the interdisciplinary analysis of a textual corpus made up by 101 end-of-degree projects. The main goal of textual analysis is the elaboration of a stylebook aimed to contribute both to the development of academic registers of Basque and to the acquisition of academic communication skills by students together with instructors in scientific and linguistic subjects. The stylebook is an essential tool for the assessment rubric to be useful and effective.

Keywords: Communication skills, assessment rubric, academic Basque speech patterns.

1 INTRODUCTION

In the European Higher Education Area (EHEA) university learning is considered as a process in which students progressively approach and are integrated into the community of experts in their disciplinary field. Communication skills are of paramount concern for this integration, since disciplinary communities are considered as epistemic and discursive communities that share goals or purposes, and use communication to achieve their goals [1]. Students acquire specialized communication patterns, in a great extent, thanks to the oral and written text input received along their learning process. Basque was introduced in higher education in the last 1970s, and nowadays is used quite extensively within university courses. In the 9 degrees offered in the Faculty of Science and Technology (Biology, Biochemistry and Molecular Biology, Biotechnology, Geology, Chemistry, Chemical Engineering Physics, Electronic Engineering, and Mathematics) 57.2 % of the students choose Basque as their main instruction language. Nevertheless, Basque language is still in the normalization process, and this process, besides a change in status, also requires the elaboration of the language for the new usages, i.e. terminological modernization and stylistic development. Specialized speech patterns must be developed and shared by experts, instructors and students, but these patterns are still unstable, changing and often far from optimal. Basque instructors at the Faculty of Science and Technology are aware of their responsibility for the progressive development of communicative competence by students. They are conscious that they have to keep in mind communication skills when they design the continuous and formative assessment of their students' progress. Nevertheless, this mission is not easy, due to the sociolinguistic situation of Basque. Just due to this situation, two elective subjects (12 ECTS) about specialized communication in Basque are offered in all degrees at the University of the Basque Country, and consequently, also in the Faculty of Science and Technology.

The end-of-degree projects elaborated in Basque constitute an interesting text genre to study the degree of elaboration of Basque academic registers, as well as a unique opportunity to evaluate the acquisition of communication skills by students at the final stage of their degree studies. In any case, both kinds of activities require an accurate distinction between, on the one hand, features or characteristics stemming from the incompleteness or/and instability of Basque academic registers and, on the other hand, deficiencies in the acquisition of these registers by some particular students. In fact,

among students and instructors that utilize Basque in the university, there are native and non-native speakers with different levels of grammatical and general communicative competence. Moreover, academic communication requires using the standard variety of the language, but standard Basque is not totally implemented, and many speakers have limited ability to use this variety. As for the end-of-degree project genre, it must be taken into account that this is one of the text varieties that might be acquired in the university [2]. Success requires learning the functional structure of the genre and the particular language patterns that are expected in this text type by assessors.

Language patterns associated with a particular text genre are also referred to as registers. Registers are described both for their linguistic features and for their situational context, since linguistic features characteristic of a register are always functional [2]. Significant and systematic distributional trends have been described across different text genres and disciplines. These trends are not arbitrary but motivated by genre-specific purposes and discipline-specific practices [3]. Register description requires analysis of relevant textual corpora, in order to identify the pervasive linguistic characteristics of this text type: linguistic features that might occur in any text type but are more common in the target register. It is very important to emphasize that the level at which we define a register depends on the aim of our analysis [2]:

There is not a correct level on which to identify a register. Rather, it depends on the goals of your study. You may want to characterize the register of academic prose, a very general register. Or you may be interested in only research articles, a more specific register within academic prose. Or you might focus on medical research articles, or even only the methods sections of experimental medical research articles. All of these can be considered registers, differing in their level of generality.

Pervasiveness of some linguistic features is the result of sharing a kind of texts by a discourse community and generating and selecting the most adequate linguistic expressions to achieve communicative goals. Despite the fact that Basque has an important presence in university teaching and learning, the use of this language in the end-of-degree projects is very recent. Consequently, the language should not have the time required to adapt functionally to the situational context, nor to fix adequate linguistic patterns and formulaic sequences. This is a very relevant fact when we approach this text type from a didactic point of view. In fact, native speakers of well-developed languages have a stock of prefabricated word chunks available to be used semi-automatically, instead of constantly making new combinations of individual words [4]. In contrast, neither Basque students aimed to write their end-of-degree projects, nor their instructors and assessors have this valuable resource, precisely because the required registers are currently under development. Furthermore, if we analyze these registers in a given stage of the development process we would detect some incorrect and/or inadequate word chunks among recurrent patterns. Functional and unsystematic speech patterns must be described and integrated in didactic tools designed to help both in the improvement of the adequacy of these patterns, and in their usage by students and instructors. We conclude that any initiative aimed to help in the development of academic registers of Basque requires a comprehensive and interdisciplinary intervention over different members of the discourse community involved in this development, i.e. instructors of scientific disciplines, instructors of language and communication, and students [5].

This work presents an interdisciplinary educational innovation project in which 13 instructors of the 9 degrees offered in the Faculty of Science and Technology at the UPV/EHU are involved together with 3 linguists, instructors of scientific communication, in that faculty. The goal of the project is the elaboration of an assessment rubric aimed to be used as a reference for students, instructors and assessors involved in the elaboration of end-of-degree projects. Nevertheless, this rubric should also be used and adapted for continuous training assessment of similar text types elaborated by students in scientific and linguistic subjects along their degree studies, such as laboratory practice's dossiers and seminar presentations. The rubric is based on the interdisciplinary analysis of a textual corpus made up by 101 end-of-degree projects. The main goal of textual analysis is the elaboration of a style book aimed to contribute both to the development of academic registers of Basque and to the acquisition of academic communication skills by students together with instructors in scientific and linguistic subjects. Consequently, the stylebook is an essential tool for the assessment rubric to be useful and effective. Since the most relevant users of both tools are not linguists, user-friendly style is required.

2 METHODOLOGY

In order to elaborate the assessment rubric and the stylebook, a corpus of 101 end-of-degree projects elaborated in Basque (654.872 text-words) was collected and linguistically annotated. The number of projects corresponding to each degree varied according to the number of students that write their project in Basque: Biology (21), Biochemistry and Molecular Biology (13), Biotechnology (7), Geology (7), Chemistry (11), Chemical Engineering (9), Physics (17), Electronic Engineering (3), and Mathematics (13). In order to facilitate the subsequent functional analysis, texts were classified according to scientific disciplines, and different sections of the end-of degree projects were processed separately. The compiled text corpus was analyzed in two different ways. On the one hand, instructors of each scientific discipline and linguists revised texts using previously agreed diagnosis criteria and categories. The aim of revision was to detect and discuss different kinds of errors and inadequacies. On the other hand, before linguistic annotation, the raw texts were also analyzed automatically using the NTLK tool to extract n-grams and frequencies. The aim of this second kind of analysis was to identify lexical bundles, which are defined as “recurrent expressions regardless of their idiomaticity, and regardless of their structural status” [6].

In order to obtain agreement on the criteria to be used in manual text analysis, all the instructors participating in the educational innovation project analyzed one and the same end-of-degree project. Instructors reflected the result of their diagnosis in a very general template with four sections: text-section, elements or features to correct or improve, description of the incorrect or inappropriate feature, explanation of the correction according to three categories (general Basque, specialized registers, and particularity of the discipline). Data collected in the first prospective diagnosis were discussed in a seminar. Taking into account the discussion by the whole team, linguists elaborated a more precise diagnosis template. Linguists explained categories included in the template to the scientists using relevant examples collected in the previous prospective diagnosis process. The diagnosis criteria and categories included in the template for the manual analysis covered all aspects of communicative competence [7]: strategic, discursive, socio-pragmatic and grammatical aspects. Seven categories were distinguished: text structure and sections, textual features (coherence, noun and verb cohesion, connection, modalization and punctuation), syntactic features, norms of the standard Basque, style and register, lexicon, and orthotypography. Experts of each disciplinary domain analyzed the end-of-degree projects of their domain and filled the corresponding template with their diagnosis results and examples. Linguists analyzed the diagnosis templates filled by scientists and concluded that, although scientists were able to detect and classify incorrect and inadequate sequences, they did not manage to distinguish between correct and incorrect uses of some structures. Consequently, a section called *zuzenketa faltsuak* ‘false corrections’ was added, in order to be taken into account when elaborating the stylebook.

As for the identification of formulaic multiword units, we confirmed the necessity for combining corpus-driven and corpus-based methodologies reported in similar works [8]. We started employing a frequency-based method to extract n-grams (2, 3, 4, 5 and 6 tokens) for identifying multiword units that occur most frequently in our corpus. Currently two parameters are used to identify lexical bundles among n-grams: frequency cut-off and dispersion (number of different texts or authors) [6]. Frequency serves to identify recurrent multiword units, and dispersion serves to avoid individual author idiosyncrasies. However, the values of these parameters are somewhat arbitrary and depend on the corpus size, and the goal and the scope of each work. Furthermore, reference works about lexical bundles are based on English corpora, and it must be taken into account that the typology of Basque language can influence dramatically those parameters’ values. Basque is a head-final free word order agglutinative language. These characteristics influence dramatically the automatic detection of recursive patterns, since some elements that often are part of English lexical bundles, such as articles and prepositions, in Basque are bound morphemes integrated in only one word or in different words. For example, the multiword unit “*the use of the*” has been described as a lexical bundle in academic writing [9]. However, the Basque equivalent is “*-aren erabilera*” in which the equivalent of the two word sequences “*of the*” are the bound suffixes *-a* and *-ren* attached to different nouns. The equivalent of the other article is the suffix *-a* attached to the noun *erabilera* ‘use’. Moreover, since the noun attached to *-aren* can vary, the frequency of a particular Basque bigram should be notably lower, compared with the English lexical bundle. The high word order freedom characteristic of Basque also affects the frequency of multiword units. For instance, the Basque equivalent of the English lexical bundle “*the goal of this work is*” should be “*lan honen helburua da*”. Nevertheless, since the verb *da* ‘is’ can be sited either in the right or in the left of its complement, order variation should notably affect the frequency of this lexical bundle.

Morphosyntactic, lexical and functional variation of lexical bundles in scientific writing has been investigated in other studies based on English corpora [10] [11]. The authors of these studies have also developed a methodological approach to address this variation. Their method is based on grouping lexical bundles by shared keywords, and applies the notion of *prototypical bundle* [10]. For instance, “*these results show / indicate / suggest that*” are an example of lexical variation obtained by grouping similar lexical bundles sharing the key word *results*. On the contrary, “*are / is / was / may be associated with*” are lexical bundles presenting morphosyntactic variation. Lexical and morphosyntactic variants can be considered as a unique speech pattern sharing the keyword *associate*. A keyword analysis evidences that “*is associated with*” is the prototypical lexical bundle, since the other variants can be explained via number, time or mood variation of the verb (*is / are / was / may be*) [11]. This method based on grouping lexical bundles and performing a key word analysis is fundamental to obtain a relevant list of Basque lexical bundles aimed at helping students in the elaboration of their end-of-degree projects. Consequently, corpus-based analysis is necessary to complement corpus-driven frequency analysis.

Finally, qualitative analysis of lexical bundles requires functional analysis and classification in functional categories. Hyland’s functional classification [9] based on a large corpus composed of research articles, Master’s theses and doctoral dissertations from four different disciplines uses three main categories and some subcategories inside them: research-oriented bundles (location, procedure bundles, quantification, description, topic); text-oriented bundles (transition signals, resultative signals, structuring signals, framing signals); and participant-oriented bundles (stance features, engagement features). Some authors [12] found the terminology of Hyland’s classification [9] difficult to understand by language learners, and adapted it using a user-friendly style for the three main categories and subcategories: describing research, organizing text, establishing stance and interacting with reader. Some adapted subcategory labels are: describing materials, expressing a change in quality, bringing the reader’s attention to a point, expressing certainty... We find this kind of functional categories very valuable for classifying lexical bundles extracted from our corpus and including them in the stylebook aimed to help students in the elaboration of their end-of-degree project.

3 RESULTS

The first stages of the manual and automatic analysis of the corpus using corpus-driven and corpus-based methodology, as well as interdisciplinary discussions in seminars, evidenced four levels to take into account, both in the elaboration of the end-of-degree projects and in their assessment:

- Structure and titles of sections.
- Correctness: orthography, syntax, lexical choices and standard Basque norms.
- Discourse organization: coherence, noun and verb cohesion, connection, modalization, punctuation, and use of fixed speech patterns
- Style trends fixed in academic prose.

In the following subsection we report a few results related with those levels.

3.1 Structure and titles of the sections

The analysis of section structure revealed important differences from one to another degree (Table 1). Only 34.7 % of the texts analysed have an abstract. Most of the abstracts are concentrated in the degree of Biology, and on the contrary, no project has an abstract in the degrees of Chemistry and Chemical Engineering. Style guidelines for the end-of-degree projects of each degree are available for students in the web of the Faculty of Science and Technology. The reason for the reported differences is that only the style guidelines for the degree in Biology specify the requirement for including an abstract. On the other hand, the major part of the projects elaborated in the degrees of Physics, Electronic Engineering and Mathematics are divided into an introduction and some sections organized according to the themes and topics. Contrary, most of the projects elaborated in other degrees describe experimental works organized just like research articles: introduction, methods, results and discussion, and conclusions. Finally, 15.8 % of the texts lack a section for conclusions.

A fact observed in some end-of-degree works is the use of inadequate titles for sections, such as “*Sarrera eta helburuak*” ‘Introduction and goals’. In our opinion, some guidelines for section structure and titles are very convenient for the stylebook, and must be taken into account in the assessment

rubric. These guidelines must describe different types of end-of-degree works (experimental, review...) and relate them to different disciplines.

Table 1. Section structure of the end-of-degree projects collected.

DEGREE	Number of Project No	Abstract	Introduction	Sections	Methods	Results	Conclusions
Biology	21	20	21	0	21	20	14
Biochemistry	13	4	12	0	13	13	11
Biotechnology	7	2	7	0	7	7	7
Geology	7	1	7	1	6	5	7
Physics	17	3	17	16	0	4	17
Electronic Engineering	3	1	2	3	0	0	2
Chemistry	11	0	11	0	11	9	11
Chemical Engineering	9	0	9	1	8	6	9
Mathematics	13	4	12	13	0	1	7
TOTAL	101	35	99	34	66	65	85

3.2 Correctness

Correct use of Basque syntax, orthography, lexical units and standard Basque norms are required for academic genres. Those elements must unquestionably be included in the assessment rubric, but only very specific cases should be explained in the stylebook.

3.2.1 Syntax

Some syntactic errors are characteristic of a low grammatical competence, for example, lack of agreement between the verb and its arguments, errors in the instantiation of the ergative suffix, and incorrect suffixes with indefinite quantifiers. Contrary, other syntactic errors are very frequent in academic texts. Consequently, far from correcting their use at de university studies, students learn these uses as register features. An example is the use of the adverbial complementizer *-elarik*, which serves to express time or cause relations, but it is found very often in academic prose attached to a subordinate clause meaning consequence. This kind of incorrect uses must be explained in the stylebook together with correct uses and relevant examples extracted from the corpus. In fact, in the diagnosis and discussion process we observed that instructors of scientific subjects find difficult to distinguish correct and incorrect uses of this kind of structures. Finally, there are particularly difficult syntactic structures characteristic of academic registers. Comparative clauses and complex noun phrases are paradigmatic examples that require main sections in the stylebook. The degree of generality and difficulty of these structures must be taken into account in the assessment rubric.

3.2.2 Standard Basque norms and orthography

It is generally accepted that academic communication requires using standard Basque. Furthermore, to a great extent, standard Basque was created precisely to facilitate the introduction of Basque in the education system as an instruction language. Consequently, we found very strange that an end-of-degree work belonging to the degree in Mathematics was written in a dialectal variety of Basque. Nevertheless, this evidence made us consider necessary to include the requirement of using standard Basque in the end-of-degree projects either in the stylebook, as well as in the assessment rubric. The standard Basque norms are published in the web of the academy of the Basque language Euskaltzaindia. A reference to this web can be included in the stylebook, but we don't find necessary to list or describe these norms.

As for orthography, students and instructors have available the automatic corrector Xuxen. The recommendation to use this tool should be included in the stylebook. On the other hand, Xuxen is a

good tool to establish a scale in the acceptability of different orthographic errors. Errors directly detected by Xuxen are less acceptable than errors in specialized lexical items that are not included in Xuxen. Homographs that are distinguished by their syntactic context and/ or semantic content require a higher competence for detecting and correcting them. Among these homographs, errors in grammar words such as *naiz / nahiz* are considered more serious than errors in lexical items such as *pauso / pausu*. Nevertheless, lexical homographs that are pervasive in Basque academic prose should be listed in the stylebook.

3.2.3 Lexical choices

It is evident that a stylebook is not a dictionary and thus an inventory of lexical elements is discarded. Nevertheless, some lexical elements are used very frequently in academic texts of almost all scientific disciplines. Some of these lexical elements are mixed up, either because their meaning are very close, or because Spanish equivalent is a polysemic word. Some paradigmatic examples are *froga / proba* 'es. prueba' 'en. evidence / test', *ikasketa / ikerketa, azterketa* 'es. estudio / estudio, investigación' 'en. study / study, reasearch', *ebazpen / bereizmen* 'es. resolución' 'en. solution, resolution / resolution (of a picture)', and *balioetsi / balioztatu* 'es. valorar' 'en. value'. Precise use of verbs is crucial in academic prose. So frequently conflictive verb uses should be listed and explained in the stylebook. Some instances of conflictive verbs and English equivalents for incorrect uses are the following: *kontsideratu* 'to consider', *suposatu* 'to suppose', *eman* 'occurr', *adierazi* 'show', *baieztatu* 'confirm', and *ondorioztatu* 'give rise to'.

3.3 Discourse organization

Coherence is a semantic-pragmatic characteristic that makes us feel that a text is meaningful. In order to achieve coherence, the author of a text must take into account the communicative situation in which the discourse is anchored, as well as the relation between the elements and pieces that make up the text. Academic discourses often require managing two different planes, i.e. description of the research and organization of the text. Students must be aware of these planes in order to success in achieving coherence. Establishing stance and interacting with reader are also important discursive functions in academic texts. Stylebook should describe this kind of discursive clues in order to improve students' discursive awareness.

Verb cohesion is the result of an adequate selection of person, time and aspect flexions depending on discursive functions. The analysis of our corpus has revealed that students often do not success to manage this matter. Consequently, some guidelines should be included in the stylebook. Guidelines should always base on functional categories and explanations.

As for noun cohesion, the analysis of our corpus showed that students often use very imprecise nominal elements, such as the demonstrative *hori* or *horrek* 'that'. Stylebook should discuss this kind of inaccuracy and provide more accurate resources, for instance, the use of generic nouns such as *parametro* 'parameter', *aldagai* 'variable', *egoera* 'state' or *prozesu* 'process' accompanying demonstratives.

As for connection and segmentation of discourse, a section of the stylebook should be dedicated to punctuation marks, since the analysis of our corpus revealed that the major part of the students do not success to use punctuation adequately. The stylebook should also offer some lists of discourse organizers, connectors and lexical bundles provided with semantic and functional information. The manual analysis of our corpus, as well as the frequency and functional analysis of lexical bundles extracted from this corpus, evidenced frequent uses of some incorrect or inadequate connectors and lexical bundles. For example, students often use the connector *hau da* 'that is' to connect a cause with its consequence. Nevertheless, the adequate use of *hau da* 'that is' is for reformulating terms or expressions in order to improve understandability. Lexical bundles should also be listed and explained in this section. Lexical bundles should be classified according to functional criteria. Under a determinate functional category, besides correct and adequate lexical bundles, such as "*lan honen helburua da*" 'the goal of this work is', "*hori dela eta*" 'as a consequence' or "*ildo honetatik*" 'in this sense', incorrect or inadequate lexical bundles obtained in the frequency-analysis should also be included: "*hau dela eta*" 'as a consequence' or "*bezala ezagutzen da*" 'is known as'.

3.4 Style trends fixed in academic prose

Finally, style trends prevalent in academic prose must be mentioned and explained in the stylebook. We include some orthotypographic uses that differentiate academic registers from other general

registers related with journalistic or literary genres. Writing of compound words, and boundaries between natural language affixes and symbols and signs belonging to other semiotic systems require their section in the stylebook.

4 CONCLUSIONS

Multidisciplinary work between instructors of scientific and linguistic subjects at the Science and Technology Faculty is required for improving academic communicative competence of students with Basque as main instruction language. The compilation and analysis of a corpus of 101 end-of-degree projects belonging to the 9 degrees offered in the Faculty, together with discussions in forums and seminars, have turn out very enriching and are making possible developing and agreeing a training assessment rubric and a stylebook. Both tools are aimed at contributing to the development of academic registers of Basque, as well as to the acquisition of academic communication skills by students and instructors.

ACKNOWLEDGEMENTS

The authors acknowledge support from the “Vicerrectorado de Innovación Compromiso Social y Acción Cultural” of the University of the Basque Country (UPV/EHU) through the SAE/HELAZ under the project “Euskarazko GRALen estilo-liburua Zientzia eta Teknologia Fakultateko titulazioetarako” (PIE5-2017/2018)

REFERENCES

- [1] J.M. Swales, *Genre Analysis. English in academic and research settings*. Cambridge: Cambridge University Press, 1990.
- [2] D. Biber & S. Conrad, *Register, Genre and Style*. Cambridge: Cambridge University Press, 2009.
- [3] N. Groom, “Pattern and meaning across genres and disciplines: An explanatory study,” *Journal of English for Academic Purposes*, vol. 4, n° 3, pp. 257-277, 2005.
- [4] J. Sinclair, *Corpus, Concordance, Collocation*. Oxford: OUP, 1991.
- [5] I. Zabala, I. San Martin, M.Lersundi & A. Elordui, “Graduate Teaching of Specialized Registers in a Language in the Normalization Process: Towards a comprehensive and Interdisciplinary Treatment of Academic Basque” in *Multiple Voices in Academic and Professional Discourse: Current Issues in Specialised Language Research, Teaching and New Technologies* (S. Maruenda-Bataller & B. Clavel-Arroitia eds.), pp.208-218, Newcastle: Cambridge Scholars, 2011.
- [6] D. Biber, S. Johansson, G. Leech, S. Conrad & E. Finegan, *Longman Grammar of Spoken and Written English*. Harlow: Person, 1999.
- [7] M. Canale & M. Swain, “Theoretical bases of communicative approaches to second language teaching and testing,” *Applied Linguistics*, vol.1, pp.1-47, 1980.
- [8] N. J. Laso & D. Salazar, “Collocations, lexical bundles and SciE-Lex” in *Biomedical English. A corpus-based approach* (I. Verdaguer, N.J. Laso & D. Salazar eds.), pp. 1-20, Amsterdam / Philadelphia: John Benjamins, 2013.
- [9] K. Hyland, “As can be seen: Lexical bundles and disciplinary variation,” *English for Specific Purposes*, vol. 27, n° 1, pp. 4-21, 2008.
- [10] D. Salazar, *Lexical bundles in Scientific English: A Corpus-based Study of Native and Non-native Writing*, Ph. D. University of Barcelona, 2011.
- [11] D. Salazar, I. Verdaguer, N. J. Laso, E. Comelles, E. Castaño & Joseph Hilferty, “Formal and functional variation of lexical bundles in biomedical English” in *Biomedical English. A corpus-based approach* (I. Verdaguer, N.J. Laso & D. Salazar eds.), pp. 39-53, Amsterdam / Philadelphia: John Benjamins, 2013.
- [12] P. Byrd & A. Coxhead, “On the other hand: Lexical bundles in academic writing and in the teaching of EAP,” *University of Sydney Papers in TESOL*, vol. 5, pp.31-64, 2010.