

DE GRUYTER

FORMULAIC LANGUAGE AND NEW DATA

THEORETICAL AND METHODOLOGICAL
IMPLICATIONS

*Edited by Elisabeth Piirainen, Natalia Filatkina,
Sören Stumpf and Christian Pfeiffer*

FORMELHAFTE SPRACHE
FORMULAIC LANGUAGE

DE
G

Formulaic Language and New Data

Formelhafte Sprache

Formulaic Language

Edited by
Natalia Filatkina, Kathrin Steyer and Sören Stumpf

Advisory Board

Harald Burger (Zurich), Joan L. Bybee (New Mexico),
Dmitrij Dobrovolskij (Moscow), Stephan Elspaß (Salzburg),
Christiane Fellbaum (Princeton), Raymond Gibbs (Santa Cruz),
Annelies Häcki Buhofer (Basel), Claudine Moulin (Trier),
Jan-Ola Östman (Helsinki), Stephan Stein (Trier),
Martin Wengeler (Trier), Alison Wray (Cardiff)

Volume 3

Formulaic Language and New Data

Theoretical and Methodological Implications

Edited by
Elisabeth Piirainen (†), Natalia Filatkina, Sören Stumpf
and Christian Pfeiffer

DE GRUYTER

The Open Access publication of the current volume was generously supported by the Interdisciplinary Research Unit “Patterns. Linguistic creativity and variation in synchrony and diachrony” established at Trier University in July 2019 within the framework of the research initiative of the Rhineland-Palatinate Ministry of Science, Professional Development and Culture.

ISBN 978-3-11-066651-9

e-ISBN (PDF) 978-3-11-066982-4

e-ISBN (EPUB) 978-3-11-066658-8



Dieses Werk ist lizenziert unter der Creative Commons Attribution-4.0 Lizenz.

Weitere Informationen finden Sie unter <http://creativecommons.org/licenses/by/4.0/>.

Library of Congress Control Number: 2019955411

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at <http://dnb.dnb.de>.

© 2020 Elisabeth Piirainen, Natalia Filatkina, Sören Stumpf and Christian Pfeiffer, published by Walter de Gruyter GmbH, Berlin/Boston

This book is published with open access at www.degruyter.com.

Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Content

Acknowledgments — v

Natalia Filatkina, Sören Stumpf & Christian Pfeiffer

Introduction: Formulaic Language and New Data — 1

Part I: Lesser-Used and Areally Limited Languages

Elisabeth Piirainen

Lesser-Used Languages and their Contribution to the Study of Formulaic and Figurative Language — 19

Stephan Elspaß

Areal Variation and Change in the Phraseology of Contemporary German — 43

Zuriñe Sanz-Villar

An Analysis of Basque Collocations Formed by Onomatopoeia and Verbs in a Translational Corpus of Literary Texts — 79

Part II: Languages Spoken outside Europe

Andreas Buerki

(How) is Formulaic Language Universal? Insights from Korean, German and English — 103

Abdullah Eisa

***Marḥaban*: Reconsidering the Criteria of an Arabic Phraseme — 135**

Muhammad A. Badarneh

Formulaic Expressions of Politeness in Jordanian Arabic Social Interactions — 151

Part III: Linguistic Varieties Used in Spoken Domains and/or Regarded as ‘Conceptually Oral’

Joanna Szerszunowicz

New Pragmatic Idioms in Polish: An Integrated Approach in Pragmateme Research — 173

Mareike Keller

Compositionality: Evidence from Code-Switching — 197

Part IV: Earlier/Historical Stages of Language Development

Marie-Luis Merten

Insights into a Changing Communal Constructicon. Legal Writing in the Late Middle Ages and Early Modern Period — 225

Christian Pfeiffer & Markus Schiegg

Religious Formulae in Historical Lower-Class Patient Letters — 249

Zuriñe Sanz-Villar

An Analysis of Basque Collocations Formed by Onomatopoeia and Verbs in a Translational Corpus of Literary Texts

Abstract: Collections of Basque proverbs and idioms have been compiled since the 16th century, but it was not until recently that interest in the study of phraseological units (PU) has arisen among researchers from different fields. This paper intends to analyze the use of a special type of phraseological unit – collocations formed by onomatopoeia and verbs – from a translational perspective in the language combination of German-Spanish-Basque. For that purpose, some introductory remarks will first be given regarding the realities of the Basque language and research on Basque phraseology. Then, the object of study will be presented, and the importance of Basque onomatopoeia will be highlighted. Since this is a corpus-based study, details on the compilation of the digitized, parallel, and multilingual corpus will be outlined, and it will be shown how these types of collocations were extracted from such a corpus in a semi-automatic way. In the translation analysis, it will be shown how one translation option stands out and how different factors (indirectness, for instance) influence Basque translations.

1 Introduction

The Basque language, spoken in the Basque Autonomous Community, Navarre, and the three provinces (Labourd, Basse-Navarre, and Soule) located within the French department of the Pyrénées-Atlantiques, is a minority language, not only due to the small number of speakers, but also because of the power relationships between the languages that coexist (i.e., the relationship between Spanish/French and Basque is not an equal, but a diglossic one), and this influences, among other things, translation activity in the Basque Country. The number of indirect translations made through Spanish versions instead of original (in this case, German) versions is, for instance, an obvious consequence, as will be shown later in this paper. At the level of language, Altzibar et al. (2011: 2) also mention that “Basque is currently undergoing a process of unification, standardisation, and adaptation to new uses. However, the influence of Spanish and

French, especially through calque, is having an almost decisive influence on this process of renovation.”

Indeed, “[i]n the last 30 years there has been an intense normalization effort in defining a unified form of Basque and in modernizing the language, as well as in extending its use from everyday affairs to high culture and science” (Uribarri 2011: 248). Although Basque has existed for centuries in many different dialects, the standard form, *Euskara Batua*, has a very short history of around 50 years, and its written literary tradition is short as well. The academic research on translation from/into Basque is very recent (Barambones Zubiria et al. 2015: 123); although authors have been compiling collections of Basque idioms and proverbs since the 16th century, not many researchers have devoted time and effort to the scientific study of phraseological Basque units.

From a contemporary and academic perspective, Aierbe (2008) analyzed the translation of phraseological units into Basque in administrative texts. An interesting conclusion she draws is the lack of fixation on the Basque language and phraseology, not only in the standard language, but also in specialized texts due to the fact that, as mentioned before, standard Basque was only recently established and because of the sociolinguistic situation of the language.

It has also been mentioned that another consequence of the diglossic situation of the Basque language is the creation and use of calques from other languages, namely Spanish and French. Some of these calques are collected on the webpage *Kalkoen Behatokia*¹ (“Observatory of Linguistic Calques”), where there is a special section on phraseological calques. During the last few years, members of this project have published papers on the use of phraseological calques in the Basque media (Alberdi et al. 2011), the use of collocations in the media (Altzibar 2004), Basque collocations (Altzibar et al. 2011), and Basque idioms (Altzibar and Bilbao 2016).

In the field of Natural Language Processing, within the IXA research group at the University of the Basque Country, authors such as Urizar (2012), Gurrutxaga (2014) and Iñurrieta et al. (2016) have worked on the automatic processing of Basque idioms and collocations.

Within Translation Studies, the author of this paper presented her PhD thesis (Sanz-Villar 2015) on the translation of phraseological units in the language combination of German-Spanish-Basque based on a parallel and multilingual corpus of literary texts. The analysis was limited to specific somatic phraseological units and binomials. The aim of the present article is to continue contributing to the field of Basque phraseological research from a translational perspective. For that

¹ <http://www.ehu.eus/en/web/eins/kalkoen-behatokia>, accessed March 28, 2018.

purpose, in the following section, the object of study will be presented: a specific type of phraseological unit, that is, Basque collocations formed by verbs and onomatopoeia, i.e. words with a more or less direct relationship between sound and meaning. Then, from a methodological perspective, the focus will be on the compilation of the corpus from which these collocations were extracted, and on the automatic extraction of these patterns. Once the collocations to be analyzed are selected and extracted, the translation analysis will be conducted and the author will conclude with some final remarks.

2 Basque Collocations Formed by Onomatopoeia and Verbs

Two approaches are usually distinguished when defining the boundaries of phraseology: the linguistic or phraseological approach and the statistical approach (Gurrutxaga 2014: 15). The former understands phraseology as a *continuum* with fixed phraseological units at the one end and flexible constructions at the other. The relation between the elements that constitute the phraseological unit is determined by their syntactic relationship and not the distance between them. According to Bernardini (2007: 1), “[p]hraseological approaches attempt to tell collocations apart from free combinations on the one hand, and from other lexical restriction phenomena on the other.” At the same time, within the statistical approach, distance or window span plays a very important role as well as the co-occurrence between the elements. As stated by Moon (1998: 26), “[c]ollocation typically denotes frequently repeated or statistically significant co-occurrences, whether or not there are any special semantic bonds between collocating items.” Word combinations are extracted from corpora and can be located in a *continuum* that goes from Sinclair’s (1987) *idiom principle* to the *open choice principle*. Here, the concept of collocation is central because collocations are more frequent in the use of language than idioms.

Following the linguistic or phraseological approach, Altzibar et al. (2011) presented a taxonomy of Basque collocations (the first and only one to the author’s knowledge) based on the well-known classification proposed by Corpas Pastor (1996) for Spanish. According to their morphosyntactic structure, Basque collocations are divided into three main classes: noun-based, adjective-based, and adverb-based. At the same time, each of them is made up of several subclasses. Among the adverb-based collocations, in the framework of this study the focus will be on one specific subclass, namely the type of collocations that are made up

of an instance of onomatopoeia (in adverbial function) and a verb, such as *dinbi-danba jo* [to hit repeatedly, to fire a shot, to toll a bell], *zanga-zanga edan* [to drink with large gulps or with great desire], or *tipi-tapa joan* [to march/walk step by step], as exemplified in Altzibar et al. (2011: 10).² A conclusion drawn by the authors of the paper indicates that “an important number of the morphosyntactic patterns of Basque collocations are different to the Romance languages of the region” (Altzibar et al. 2011: 11). The collocation type “onomatopoeia + verb” is mentioned as one of those different patterns. Apart from this feature, it is worth mentioning that there is a large number of them in not only everyday but also educated language.

According to Ibarretxe-Antuñano (2006), the linguistic study of onomatopoeia has long been neglected. However, studies on onomatopoeia carried out in different languages have shown that they constitute an essential part of a language and that they are of considerable importance in terms of quantity.³ Basque onomatopoeia are characterized by three features: total or partial reduplication, the use of unusual phonological and prosodic elements (for instance, *-dz*), and the association of certain sounds with certain meanings (Ibarretxe-Antuñano 2006: 151). According to Ibarretxe-Antuñano, the former feature does not appear very often in other European languages, but it is one of the most used strategies in Basque (not only with regard to onomatopoeia but also as a mechanism to express emphasis).

From a morphosyntactic point of view, different grammatical categories can be found (nouns, adverbs, verbs, adjectives, interjections) among Basque onomatopoeia, and, due to the fact that in Basque it is quite easy to create new words through derivation and composition, it is not unusual for new words to be created based on onomatopoeia (Ibarretxe-Antuñano 2006: 153). Semantically, they are mainly used in the following semantic fields: actions and activities, animals, plants, atmospheric phenomena, musical instruments, physical and mental features, tools, things, child language, large quantities, nature, and sexual terms. The first group is the largest among Basque onomatopoeia, and is divided into

² Ibarretxe-Antuñano (2006: 153), making reference to Etxepare (2003), refers to this type of collocation as “complex predicates” and adds that they are very common in languages with onomatopoeia. She also mentions that the verbs accompanying the onomatopoeia are usually “dummy verbs”, i.e., verbs such as *(to) make*, *(to) say*, *(to) think*, and so on. Therefore, it is not the verb but the onomatopoeia that provides the real meaning of the construction (Ibarretxe-Antuñano 2012: 150–151).

³ Referring to Basque, Schuchardt (1925: 18), for instance, says that “das Baskische ist sehr reich an deutlichen Schallworten”, emphasizing the abundance of onomatopoetic constructions in Basque.

different categories: motion, communication, light, sound, beverage/food, destruction, hitting, boiling, emotions, body functions, and others (Ibarretxe-Antuñano 2012: 152–160).

The main limitation of Ibarretxe-Antuñano and Martínez Lizarduikoa (2006) is that it does not reflect the real use of these constructions, the onomatopoeia, since all the examples are extracted from different dictionaries and collections containing Basque multiword expressions.⁴ Ibarretxe-Antuñano also concludes that studies regarding the translation of onomatopoeia constitute a field that requires further research in the future (Ibarretxe-Antuñano 2012: 171). This paper intends to make a contribution to filling this gap by analyzing the use of collocations formed by onomatopoeia in children's and adult literature texts translated from German into Basque.

3 Compilation and Features of the Corpus

In order to conduct such an analysis, it was first necessary to compile a corpus consisting of German-into-Basque literary translations. For that purpose, the following steps were taken: the creation and description of a catalog consisting of literary texts that have been translated in the language combination German-Basque, specification of the criteria for selection of the texts that would be part of the corpus, and (once the texts were selected) digitization, cleaning, tagging, aligning, and uploading the texts to a database.

The catalog, known as Aleuska, which was updated until 2013, includes 710 entries that can be divided into different text types: theater (1%), essay (8%), adult literature (8%), poetry⁵ (20%), and children's literature (63%). The first decision in the selection of the texts was based on this distribution: Children's literature (CL) texts and adult literature (AL) texts were included in the corpus on the basis of their representativeness in the catalog. The chronological factor was another aspect that was taken into account when defining the selection criteria. In the catalog, as far as AL and CL texts are concerned, the number of translations published starts increasing from the year 1980 onwards. For this reason, the texts included in the corpus are translations published from that year on. In addition,

⁴ However, in the paper from 2012, Ibarretxe-Antuñano includes a section about the pragmatic functions of Basque onomatopoeia, and in this section the reader can find how these constructions are used in different contexts, such as oral literature, poetry, comics, and advertising.

⁵ Regarding poetry, it has to be mentioned that many entries in the Aleuska catalog consist of single poems.

the CL-AL sub-catalog includes texts by 126 different German authors and 127 different translators. Thus, source and target author diversity was another criterion to be considered when selecting the texts. The final factor that needs to be mentioned is the mode of translation; in other words, the fact that the CL and AL texts of the Aleuska catalog were identified as direct translations made from the German source text or as indirect translations carried out from an intermediary text (most of the time the Spanish version) was of great importance (not only for the compilation of the corpus but also for the translation analysis).

To sum up, the corpus is made up of 24 CL texts and 24 AL texts, which represents around 3.5 million words. As for author diversity, works from 30 different German authors and 28 different translators were selected; regarding the mode of translation, there are 34 direct translations and 14 indirect translations. With reference to this last feature, it is important to bear in mind that they should be regarded as *assumed* direct and indirect translations, since it was not always easy to obtain this information from the catalog.

For the compilation of the corpus, a tool called TAligner was used. The latest versions of this program were developed within the TRALIMA/ITZULIK research group⁶ at the University of the Basque Country, and one of its strong points is that it allows the simultaneous alignment of not only two but various texts. This option was an indispensable condition for the compilation of the present corpus—indispensable in order to be able to align the indirect translations with their source and intermediary texts.

After all the bi- and tri-texts had been digitized, they were cleaned, tagged, and aligned. It can clearly be observed in Figure 1⁷ that the tool contains all these functions—*limpiar* (clean), *etiquetar* (tag), and *alinear* (align)—and that it allows the simultaneous alignments of numerous texts, three in this case:

⁶ The website of the research group can be accessed via this link: <https://www.ehu.eus/en/web/tralimaitzulik/home>, accessed March 28, 2018.

⁷ Figures 1 and 2 represent the latest version of the tool, known as TAligner 3.0.



Fig. 1: Functions of the program TAligner 3.0

Once all the texts are aligned, the user can make queries by using the option *Consultar Corpus*. The example in figure 2, for instance, shows the search for the Basque collocation *kar-kar egin* [to laugh] in the corpus. In this case, the author was querying the whole corpus, but since metadata, such as author’s name, translator, genre, and so on, was introduced when tagging the texts, the queries can be limited to words or word combinations in certain texts translated by certain translators or texts written by certain authors, just to mention some possibilities.

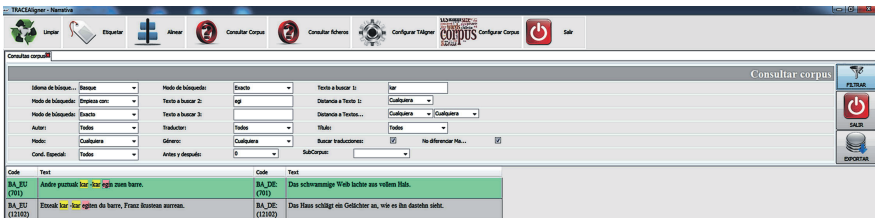


Fig. 2: Querying the corpus with the program TAligner 3.0

4 Extraction of Potential Collocations

The aim is to extract specific grammatical patterns from the corpus that may constitute collocations. For this purpose, it is necessary to tag the corpus at a part-of-speech (POS) level. This task was performed with the help of a Natural Language

Processing tool called IXA pipes⁸ and developed at the University of the Basque Country by members of the IXA research group (Agerri et al. 2014). Although this tool allows texts to be tagged at different levels, the annotation options used in the present case were tokenization, lemmatization, and POS tagging. As for the process, only the TXT files of each text need to be annotated—the Basque target texts in this case—and the tool will provide the user with an output file in NAF format for each file that is tokenized, lemmatized, and tagged at the POS level.

To conduct the extraction of the potential collocations, a toolkit called Foma (Hulden 2009) was employed. Created by Mans Hulden, it is a free and open-source tool, and may be used to satisfy different goals within Natural Language Processing. For the purposes of the present study, a code to extract user-defined grammatical patterns based on the information of the NAF tagged files was defined⁹ and then processed with Foma.

The number of patterns extracted was very large (48,296), a result which was expected considering the widespread use of these patterns as well as the size of the corpus. However, advantage was taken of a special feature of the Basque onomatopoeia under study: The fact that they contain a hyphen in the onomatopoeia part of the word combination helped reduce the initial list to 2,617 patterns. Then, with the help of a number of Unix commands (*sed*, *egrep*, and so on) and some manual work, unnecessary combinations were removed¹⁰ and a list of 428 patterns was compiled. Many of those combinations were partial or total reduplications that do not constitute onomatopoeia. As mentioned in Ibarretxe-Antuñano (2012: 138), reduplications, constituting onomatopoeia or not, are a frequently employed resource in Basque. From the 428 patterns, 147 different types were manually identified as potential collocations.

Since many of those 147 patterns occur in the corpus several times, it was decided to limit the study to those patterns that, according to the semantic classification mentioned in section 2, represent the largest group among Basque onomatopoeia: the ones that describe the semantic field of actions and activities. All in all, 66 types and 162 tokens were selected and queried for the subsequent translation analysis.

⁸ <http://ixa2.si.ehu.es/ixa-pipes/>, accessed March 28, 2018.

⁹ I hereby thank the member of the IXA research group Iñaki Alegria for helping me to write the code.

¹⁰ Many of them constituted compounds, such as *gutun-azala ireki* [to open the envelope] and were easy to identify and exclude from potential collocations. Many other results indicated that unnecessary spaces were added between the constituents of a compound, and thus, only a part of the compound was extracted. For instance, in *monopoly-jolastu* [to play Monopoly], the word *game* or something similar is missing.

The ascription of semantic fields to the patterns found in the corpus was not always straightforward. While in some cases there was no doubt about the category a specific collocation belonged to (for instance, *mauka-mauka jan*, *dir-dir egin*¹¹), other cases (such as, *plisti-plasta egin*¹²) were more controversial. Be that as it may, it is worth mentioning that the largest semantic field, in terms of occurrences, was the field expressing motion, with 44 occurrences. In fact, as for the group of onomatopoeia describing motion, Ibarretxe-Antuñano (2012: 153) mentions that it is one of the largest in Basque: “El grupo de las onomatopeyas que describen el movimiento es uno de los más numerosos en euskera.”

Analyzing the verbs that accompany the onomatopoeia, it may be concluded that the most usual by far is the above-mentioned “dummy verb” *egin* [to do; to make] with 88 occurrences, followed by the verb *joan* [to go] with 14 occurrences. Thus, in the case of those dummy verbs, as mentioned in Ibarretxe-Antuñano (2012: 150–151), the onomatopoeia is the component providing the real meaning of the construction.

5 Translation Analysis

During the extraction of the collocations and the translation analysis, the main focus was on the target texts and culture, which is in line with the target-oriented approach established within Descriptive Translation Studies (Toury 2012). In other words, the queries were made based on the Basque target texts, and thus the Basque translations and their cultural context served as the starting point. Then, the target texts’ outputs were compared to their German (and Spanish) source texts, and the following translation options were identified:

- *No PU-Collocation*: The counterpart of the Basque collocation is not a PU in the German source text; it is usually a single verb.
- *Collocation-Collocation*: A German collocation has been translated with another collocation in the Basque target text.
- *Idiom-Collocation*: The German equivalent of the Basque collocation is an idiom.

¹¹ *Mauka-mauka jan* means ‘to eat voraciously, greedily’ and *dir-dir egin* ‘to shine; to sparkle; to gleam; to glitter’, and they clearly belong to the semantic groups *beverage/food* and *light*.

¹² In *plisti-plasta egin*, the word *plisti-plasta* is the onomatopoeia of splashing water. Due to the sound created through the action of splashing water, it was ascribed to the semantic group of *sound*, but it may also belong to the semantic field expressing motion.

- *Ø-Collocation*: There is no counterpart in the German source text for the Basque collocation; the Basque collocation is an addition.

The examples in table 1 may serve for a better understanding of each of the translation options explained above:

Tab. 1: Examples of the different translation options found in the corpus

Source Text	Target Text	Translation Option
1. Ein gigantischer Fluss schlängelt sich behäbig seinen Weg. (HL)de ¹³ [A gigantic river meanders its way slowly. ¹⁴]	Ibai ikaragarri handi batek bere bideari jarraitzen dio astiro-astiro, sigi-saga eginez . (HL)eu ¹⁵ [A gigantic river follows its way slowly, zig-zagging.]	No PU-Collocation
2. [...] Franz geht mit kleinen Schritten und weiß: (BAde) ¹⁶ [Franz walks in small steps and knows:]	[...] Franz tipi-tapa doa eta badaki: (BAeu) ¹⁷ [Franz walks pitter-patter and knows:]	Collocation-Collocation
3. Das schwammige Weib lachte aus vollem Hals . (BAde) [The spongy woman laughed loudly.]	Andre puztuak kar-kar egin zuen barre . (BAeu) [The swollen woman laughed loudly.]	Idiom-Collocation
4. Es gibt auch unter uns Zöglingen etwas wie einen aus Luft und Nichts herausgegriffenen Zeitungenklatsch. (JGde) ¹⁸ [There is also something like newspaper gossip plucked out of the air and nothing among us pupils.]	Ikasleone artean ere badabil zerbait, airetik eta hutsetik bolo-bolo zabaltzen diren egunkarietako hizkimizkien antzekorik. (JGeu) ¹⁹ [There is also something among us pupils, similar to the newspaper gossip that extends everywhere, plucked out of the air and nothing.]	Ø-Collocation

¹³ Knister (2003): *Hexe Lilli auf der Jagd nach dem verlorenen Schatz*. Würzburg: Arena.

¹⁴ All English translations are by the author of the paper. They are intended to show the literal meaning of the examples.

¹⁵ Knister (2003): *Kika Supersorgina altxorraren bila*. Bilbao: Gero (translator: Iñaki Aristondo).

¹⁶ Döblin, Alfred (1929): *Berlin Alexanderplatz*. Berlin: Fischer.

¹⁷ Döblin, Alfred (1929): *Berlin Alexanderplatz*. Berlin: Fischer (translator: Anton Garikano).

¹⁸ Walser, Robert (1909): *Jakob von Gunten*. Berlin: Cassirer.

¹⁹ Walser, Robert (2005): *Jakob von Gunten*. Donostia: Erein (translator: Edorta Matauko).

The first example shows how the equivalent of the German verb *sich schlängeln* ('to meander') and the Spanish verb *serpentea* is a collocation made up of the onomatopoeia *sigi-saga* [zigzag] and the verb *egin* [to make; to do]. The fact that a reduplication (*astiro-astiro*, 'slowly') appears in the same sentence in Basque makes this example even more interesting because it shows (together with other examples that were found in the corpus but cannot be mentioned in this paper due to space constraints) that not only onomatopoeia but also reduplications are a widespread resource in Basque, as mentioned in section 2. The translation of the German collocation *mit kleinen Schritten gehen* [to walk in small steps], as identified in the online collocation dictionary compiled by Häcki Buhofer et al. (2014)²⁰, is the Basque collocation *tipi-tapa joan* [to walk pitter-patter]. In the third example, the German idiom *aus vollem Hals* [loudly] is translated with the Basque collocations *kar-kar barre egin* [to laugh loudly]. In the last case, it can be observed that there is no counterpart for the Basque onomatopoeia *bolo-bolo* [everywhere] and that it rather serves to intensify the meaning of the two phraseological units *aus dem Nichts* [out of nothing] and *aus der Luft gegriffen sein* [something that is plucked out of the air].

The distribution of the above-mentioned translation options across the different subcorpora—children's literature direct translations (CL DI), adult literature direct translations (AL DI), children's literature indirect translations (CL INDI), and adult literature indirect translations (AL INDI)—is presented in tables 2 and 3. Before giving any interpretation of the figures, it is important to mention that the different subcorpora are not equal in size. The largest is AL DI (1,708,825 words), followed by CL DI (809,301 words); AL INDI (547,773 words) and CL INDI (463,634 words) are significantly smaller compared to the former. Therefore, it cannot be concluded that more collocations formed by onomatopoeia and a verb have been extracted from the AL DI subcorpus. Rather, the main conclusion that can be drawn from the figures in tables 2 and 3 is the predominance of the translation option *No PU-Collocation* in all subcorpora, and consequently the sporadic occurrence of the rest of the translation options.

²⁰ <http://www.kollokationenwoerterbuch.ch/web/>, accessed March 28, 2018.

Tab. 2: Distribution of translation options across subcorpora (raw numbers)

	CL DI	AL DI	CL INDI	AL INDI	Total
No PU-Collocation	46	72	16	10	144
Collocation-Collocation	2	6	1	0	9
Idiom-Collocation	1	4	1	1	7
∅-Collocation	0	1	0	1	2
	49	83	18	12	162

Tab. 3: Distribution of translation options across subcorpora (in percentages)

	CL DI	AL DI	CL INDI	AL INDI	Total
No PU-Collocation	93.88	86.75	88.89	83.33	88.89
Collocation-Collocation	4.08	7.23	5.56	0.00	5.56
Idiom-Collocation	2.04	4.82	5.56	8.33	4.32
∅-Collocation	0.00	1.20	0.00	8.33	1.23
	100.00	100.00	100.00	100.00	100.00

Next, it is intended to go beyond these general figures and focus on the nuances of the different translation options as well as other interesting features observed during the translation analysis.

5.1 No PU-Collocation

Despite the undeniable predominance of this translation option, it is necessary to mention that it is not a homogeneous group and that it deserves more detailed attention. In some cases, for instance, there is no phraseological unit in the German source text, but instead verbs with a more or less evident onomatopoeic origin can be found, such as *schwabbeln* [to wobble], *plappern* [to chatter], *kitzeln* [to tickle], *plantschen* [to splash], *scheppern* [to rattle], *brummen* [to grumble], *knurren* [to growl], and so on. In the first example, extracted from an adult literature text, to express that Mieke is chattering, in the original text the verb *plappern* is used, and in the Basque text the onomatopoeia *tar-tar*, together with the dummy verb *esan* [to say], is employed. The second example is from a children's

literature text, and the Basque onomatopoeia *mar-mar egin* was selected as a counterpart of the German onomatopoeic verb *brummen*.

Tab. 4: The use of verbs of onomatopoeic origin in the source texts

Source Text	Target Text
1. Und Mieke sitzt auf, faßt ihren Franz um und sieht ihm wonnig ins Gesicht und plappert so lauter süßen Quatsch und bettelt und bettelt: (BAde) [And Mieke sits up, takes her Franz and looks him in the face with delight and chatters, saying nothing but sweet nonsense and begs and begs.]	Eta Mizek bizkarra tentetu, besarkatu bere Franz eta begirutzen dio aurpegira bozkariotsu eta tar-tar esaten dizkio sekulako txorakeria goxoak eta arren eta arren: (BAeu) [And Mieke stiffens her back, hugs her Franz and looks at his face with joy and does not stop telling him some incredible sweet nonsense and begs and begs.]
2. “Ja, ja”, brumnte Tobi, der sich mindestens fünf Jahre zu alt für eine Holzseisenbahn fühlte. (SLHde) ²¹ [“Yes, yes,” grumbled Tobi, who felt at least five years too old for a wooden train.]	“Bai, bai”, mar-mar egin zuen Tobik, zeini zurezko trenekin jostatzeko gutxienez bost urte zaharregia zela iruditzen zitzaion. (SLHeu) ²² [“Yes, yes,” grumbled Tobi, who felt at least five years too old to play with a wooden train.]

On other occasions, there is not just one verb, but two (of onomatopoeic or non-onomatopoeic origin) in the German source text, and the translators decided to use a collocation with onomatopoeia in the Basque target texts. Two different examples can be found in table 5: In the first example, the meaning of the German verbs *rappeln und rattern* [to rattle] is represented by the Basque word combination *triki-traka ibili* [to clatter], while in the second example, the above-mentioned verb *plappern* is repeated twice. The Basque translator decided to triplicate²³ the onomatopoeia by using *tar-tar-tar*, thus emphasizing that she does nothing but chatter.

²¹ Sommer-Bodenburg, Angela (1991): *Schokolowski. Lustig ist das Hundeleben*. Munich: Bertelsmann.

²² Sommer-Bodenburg, Angela (1996): *Txokoloski. Dibertigarria da txakurren bizitza*. Bilbao: Desclée de Brouwer (translator: Edurne Azkue).

²³ Ibarretxe Antuñano (2012: 138) refers to this structural feature as *triplicación total* or total triplication.

Tab. 5: The use of two verbs of onomatopoeic origin in the source texts

Source Text	Target Text
1. Neben ihm rappelte und ratterte schon der Kompressor. (JVde) ²⁴ [Beside him, the compressor was already rattling.]	Ondoan konpresorea triki-traka zarataka zebilen jada. (JVeU) ²⁵ [Beside the compressor was already clattering.]
2. Sie plappert, plappert , legt den Kopf um seinen Hals, [...]. (BAde) [She chatters, chatters, puts her head around his neck.]	Tar-tar-tar ari da, jarri du burua mutilaren lepoan, [...]. (BAeu) [She does not stop talking, puts her head on the boy's neck.]

In many cases, the most natural, direct, and straightforward correspondent of the German verb is the collocation with onomatopoeia, but in other cases, the Basque translation option seems more phraseological or expressive than the original. As can be seen in table 6, the first example is extracted from an indirect translation. Both the German and Spanish texts contain a single verb (*trinken*, *beber*, 'to drink'), while the translator in the Basque version, with the use of the onomatopoeia, describes more exhaustively how the drinking is performed: *zanga-zanga* [in gulps]. A very similar situation can be observed in the second example: The equivalent for the German verb *essen* [to eat] is not just the Basque verb *jan*, but it describes how the fish is being eaten by Jasper: *mauka-mauka* [voraciously, greedily].

Tab. 6: Examples of more phraseological or expressive options in the target texts

Source Text	Intermediary Text	Target Text
1. Er schimpft, reißt den Kühlschrank auf, schnappt sich das Mineralwasser und trinkt direkt aus der Flasche. (HL)de [He scolds, tears open the fridge, grabs the mineral water and drinks directly from the bottle.]	Dani abre la nevera de golpe y bebe directamente de la botella de agua mineral. (HL)es [Dani opens the fridge at once and drinks directly from the bottle of mineral water.]	Danik hozkailua kolpetik ireki eta ur minerala zanga-zanga edaten du botilatik bertatik. (HL)eu [Dani opens the fridge at once and drinks in gulps the mineral water directly from the bottle.]

²⁴ Massanek, Joachim (2003): *Juli, die Viererkette*. Frankfurt am Main: Baumhaus.

²⁵ Massanek, Joachim (2010): *Juli, defensa onena*. Bilbao: Gero (translator: Nuria Sebrango).

Source Text	Intermediary Text	Target Text
2. Jasper fanden wir bei Mama in der Küche. Er aß den gestern verschmähten Tiefkühlfisch. (DAKde) ²⁶ [We found Jasper in the kitchen with mum. He ate the yesterday spurned frozen fish.]		Jasper, amarekin sukaldean aurkitu genuen, bezperan nahi izan ez zuen arraina maukamauka jaten . (DAKeu) ²⁷ [We found Jasper in the kitchen with mum, eating voraciously the fish he did not want yesterday.]

The following translation options in the Basque target texts extracted from indirect translations could be described as more phraseological if the original German text and the Basque translation were compared, but it is obvious from looking at the intermediary versions that the Basque translation has been influenced by the Spanish version. In the first example, the sun rests [*ruhen*] in the German text, while in the Spanish and Basque versions it sparkles [*centellear, diz-diz egin*]. In the second example, the path leads [*führen*] somewhere in the German original text, and in the Spanish and Basque versions it meanders or zigzags [*serpentear, sigi-sagan igaro*].

Tab. 7: Examples of the influence of the intermediary texts on the Basque translations

Source Text	Intermediary Text	Target Text
1. Wenn das liebe Tal um mich dampft, und die hohe Sonne an der Oberfläche der undurchdringlichen Finsternis meines Waldes ruht , und nur einzelne Strahlen sich in das innere Heiligtum stehlen [...]. (DLWde) ²⁸ [When the dear valley steams around me, and the high sun rests on the surface of the im-	Quando el valle se vela en torno mío con un encaje de vapores; cuando el sol de mediodía centellea sobre la impenetrable sombra de mi bosque sin conseguir otra cosa que filtrar entre las hojas algunos rayos hasta el fondo del santuario;	Harana nire gainean lur-rinezko mihiztaduraz bar-randatzen denean; eguerdiko eguzkiak nire basoaren ger-izpe sarkaitzaren gainean, hostoen artean eta santutegiko ostalderaino izpi batzu iraitzea baino lortu ez dela, diz-diz egiten duenean;

²⁶ Nöstlinger, Christine (1982): *Das Austauschkind*. Weinheim: Beltz & Gelberg.

²⁷ Nöstlinger, Christine (1991): *Ingeles bat etxean*. Lizarra: Elkar (translator: Xabier Mendiguren).

²⁸ Goethe, Johann Wolfgang von (1774): *Die Leiden des jungen Werther*. Leipzig: Weygand.

Source Text	Intermediary Text	Target Text
penetrable darkness of my forest, and only single rays steal into the inner sanctum.]	(DLWes) ²⁹ [When the valley veils around me with a lace of vapors; when the midday sun sparkles on the impenetrable shadow of my forest, achieving nothing but filtering some rays through the leaves to the bottom of the sanctuary.]	(DLWeu) ³⁰ [When the valley lurks over me with a lace of steam; when the midday sun sparkles on the impenetrable shadow of my forest, achieving nothing but filtering some rays through the leaves and to the sanctuary's heaven.]
2. Dieser Weg, der direkte Weg nach Napoule, führte an den Ausläufern des Tanneron entlang durch die Flußsenken von Frayere und Siagne. (DPGMde) ³¹ [This path, the direct path to Napoule, led along the foothills of the Tanneron through the river valleys of the Frayere and the Siagne.]	Este camino, el camino directo a Napoule, serpenteaba por las estribaciones del Tanneron, cruzando las cuencas de Frayère y Siagne. (DPGMes) ³² [This path, the direct path to Napoule, meandered through the foothills of the Tanneron, through the basins of the Frayere and the Siagne.]	Bide hura, La Napoulera zuzenean zihoan bidea, sigisagan igarotzen zen Tanneronen oinetik, Frayere eta Siagne ibaien arroetan barrena. (DPGMeu) ³³ [This path, the direct path to Napoule, zigzagged through the foothills of the Tanneron, through the basins of the Frayere and Siagne rivers.]

5.2 Collocation-Collocation

A special case of this translation option is represented by the example in table 8. The Basque children's literature text has been translated indirectly through the Spanish text. However, in this case, the collocation is found in the German text (*laut klopfen*, 'to knock loudly') and not in the Spanish text, where a single verb has been used (*dar*, 'to knock'). Thus, given the indirect character of the Basque translation, it may be argued that, with the use of the word combination *kaska-kaska jo* [to tap] as an equivalent of the Spanish verb *dar*, a more phraseological target text is created.

²⁹ Goethe, Johann Wolfgang von (1835): *Las desventuras del joven Werther*. Barcelona: Apolo (translator: José -Mor de Fuentes).

³⁰ Goethe, Johann Wolfgang von (1987): *Werther*. Donostia: Kriselu (translator: Gotzon Lobera).

³¹ Süskind, Patrick (1985): *Das Parfum*. Zürich: Diogenes.

³² Süskind, Patrick (1985): *El perfume*. Barcelona: Círculo de lectores (translator: Pilar Giralte).

³³ Süskind, Patrick (2007): *Perfumea*. Irun: Alberdania (translator: Miren Arratibel).

Tab. 8: Example of the translation option *Collocation-Collocation* in an indirect translation

Source Text	Intermediary Text	Target Text
denn ein Gast klopft laut mit dem Messer ans Glas und will zahlen. (EDde) ³⁴ [because a guest knocks loudly with the knife on his glass and wants to pay.]	porque un parroquiano daba con el cuchillo en la copa como el que quiere pagar. (EDes) [because a customer knocks with the knife on his glass as if he wants to pay.]	bezero bat, ordaindu nahi zue-narena eginez, aiztoarekin basoa jotzen ari bait zen kaskakaska . (EDeu) [because a customer, as if he wanted to pay, was tapping his glass with the knife.]

5.3 Idiom-Collocation

In the example cited in table 9, which represents an indirect translation, the idiom is also found in the German source text (*wie Kraut und Rüben liegen*, ‘to be higgledy-piggledy’), and not in the intermediary text from which the Basque translation originates. However, since the meaning changes in the Spanish version [to curl up] with respect to the original version, the Basque (which matches the meaning of the Spanish text) and German versions also differ semantically.

Tab. 9: Example of the translation option *Idiom-Collocation* with a difference in meaning due to indirectness

Source Text	Intermediary Text	Target Text
Die Jungen warfen sich zu Boden und lagen wie Kraut und Rüben durcheinander. (EDde) [The boys threw themselves to the ground and lay higgledy-piggledy.]	Y los chicos se tiraron al suelo y se quedaron muy acurrucaditos . (EDes) ³⁵ [And the boys threw themselves to the ground and they were very curled up.]	Mutikoak etzan eta kuzkur-kuzkur eginda geratu ziren. (EDeu) ³⁶ [The boys lay down and they were very curled up.]

In the next and final example, however, idioms can be observed both in the original and in the intermediary text: *mit vollen Backen (verzehren)* (‘to eat with

³⁴ Kästner, Erich (1929): *Emil und die Detektive*. Zurich: Atrium.

³⁵ Kästner, Erich (1967): *Emilio y los detectives*. Barcelona: Juventud (translator: José Fernández).

³⁶ Kästner, Erich (1991): *Emilio eta detektibeak*. Donostia: Elkar (translator: Tomás Sarasola).

stuffed cheeks’) and (*devorar*) *a dos carrillos* (‘gobble food down’). The Basque translator used a collocation formed by onomatopoeia and a verb (*mauka-mauka jan*, ‘to eat voraciously, greedily’) to represent the same meaning.

Tab. 10: Example of the translation option *Idiom-Collocation* with no difference in meaning

Source Text	Intermediary Text	Target Text
[...] und, wenn sie das gewünschte endlich erhaschen, es mit vollen Backen verzehren und rufen: “mehr!” (DLWde) [and when they finally catch what they desire, eat it with stuffed cheeks and shout: “more!”.]	[...] cuando logran atrapar el manjar apetecido lo devoran a dos carrillos y gritan: “¡Más!” (DLWes) [and when they finally catch the desired delicacy, they gobble it down and shout: “more!”.]	[...] gura izandako jatena harrapatzen lor dezatenean mauka-mauka jan eta “gehiago” deiadar egiten dutenak direla. (DLWeu) [when they finally catch the desired delicacy, they are the ones who eat it voraciously and shout “more”.]

6 Conclusions

This paper has focused on specific phraseological units found in the Basque language. Basque has a weak tradition of written literature and a short history of its standard variety, which coexists diglossically with other major languages and is still in the process of standardization. Given these features, it is understandable that there has been no systematic research in the field of Basque phraseology and that even less attention has been paid to the study of the translation of phraseological units from/into Basque. However, as presented in the introduction of this paper, there are some projects worth mentioning (Aierbe 2008; Sanz-Villar 2015; Iñurrieta et al. 2016), and the objective of the present article has been to make a further contribution toward research in this field.

Altzibar et al. (2011) identified collocations formed by partially or totally duplicated onomatopoeia and a verb as a special type of formulaic pattern in Basque, and the actual use of those collocations has been analyzed in this paper from a translational perspective. By doing so, we have encountered both theoretical and methodological challenges since there was no similar previous study that would have served as a reference. From a theoretical perspective, it has not always been easy to identify the onomatopoeia part of the collocation as actual onomatopoeia or to establish the boundaries between collocations and free word combinations. In this sense, the onomatopoeia dictionary of Ibarretxe-Antuñano

and Martínez Lizarduikoa (2006) as well as Ibarretxe-Antuñano's papers (2006, 2012) on this topic were of great help and served as a constant reference.

Methodologically, the first challenge was to create a corpus from scratch that would meet the requirements to conduct German-into-Basque translation analyses of literary texts. Although at present there are a number of very diverse Basque corpora available to any user, the type of corpus we needed – a digitized, parallel, and multilingual corpus consisting of German original texts, intermediary versions in the case of indirect translations, and Basque target texts – not only needed to be built from scratch but also required creation of a specific tool,³⁷ which allowed for the simultaneous alignment of several texts. Due to the lack of precedent, another methodological challenge involved the extraction of the phraseological units under analysis. This was solved thanks to a lemmatization and POS-tagging tool developed by members of the IXA group that works, among others, with Basque, and an open-source tool that requires some knowledge of computational linguistics, but can be used with any language. The author is aware of the fact that the corpus probably contains more collocations formed by onomatopoeia that were not extracted using this method, because, for instance, the verb does not always appear right after the onomatopoeia. However, a great number of them were extracted, and this may be seen as a first attempt toward such an exhaustive analysis.

The translation analysis has shown that, despite the predominance of the translation option *No PU-Collocation*, the nuances that are hidden behind it are of great significance from a translational point of view. Sometimes the use of the collocations seemed to be the most natural way of rendering the content of the source text in the target text, but at other times, as exemplified in table 6, the use of the Basque collocations resulted in a more phraseological text. When analyzing somatic phraseological units (Sanz-Villar 2018), it was concluded that in the case of indirect translations there is “a tendency to deviate from the Spanish intermediary version and create more typical Basque texts.” The examples in tables 6 and 8, as well as others that could not be presented in this paper, corroborate this hypothesis, but more examples should be analyzed. One thing is clear: during the whole process, from the creation of the catalog to the translation analysis, it was crucial to take into account that indirect translations are a reality in German-into-Basque translations.

³⁷ I hereby thank the computer technician Iñaki Albisua for his dedication to the alignment tool over the past years.

References

- Aggeri, Rodrigo, Josu Bermudez & German Rigau (2014): IXA pipeline: Efficient and ready to use multilingual NLP tools. *Proceedings of the 9th Language Resources and Evaluation Conference (LREC2014)*, 26–31.
- Aierbe, Axun (2008): La traducción a la lengua vasca de las unidades fraseológicas especializadas del lenguaje administrativo. In María Isabel González (ed.), *A Multilingual Focus on Contrastive Phraseology*, 27–44. Hamburg: Dr. Kovač.
- Alberdi, Xabier, Xabier Altzibar & Julio Garcia (2011): Calcos fraseológicos en euskera de los medios de comunicación. In Antonio Pamies, Lucía Luque Nadal & José Manuel Pazos Breña (eds.), *Phraseologie und Parömiologie. Multi-Lingual Phraseography: Second Language Learning and Translation Applications*, 215–224. Baltmannsweiler: Schneider Verlag Hohengehren.
- Altzibar, Xabier (2004): Kolokazioak euskaraz. Zer axola duten kazetaritzan. *Euskarazko kazetaritzaren I. Kongresua. Kazetaritza Euskaraz: Oraina eta Geroa*, 383–395.
- Altzibar, Xabier & Xabier Bilbao (2016): Locuciones en euskera: Necesidad y pautas para su recopilación y ordenación a partir de los corpus textuales existentes. In Gloria Corpas Pastor (ed.), *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives*, 24–31. Geneva: Editions Tradulex.
- Altzibar, Xabier, Xabier Bilbao & Koldo Garai (2011): Collocations in Basque: A test for classification. *Proceedings of the 5th International Conference on Meaning-Text Theory*, 1–12.
- Barambones Zubiria, Josu, Elizabete Manterola Agirrezabalaga, Zuriñe Sanz Villar, Ibon Uribarri Zenekorta & Naroa Zubillaga Gómez (2015): Itzulpen ikasketak eta euskara. Zenbait ekarpen Itzulpen Ikasketak Deskribatzaileei, eta haratago. In José Jorge Amigo Extremera (ed.), *Traducimos Desde el Sur. Actas del VI Congreso Internacional de la Asociación Ibérica de Estudios de Traducción e Interpretación*, 121–140. Las Palmas de Gran Canaria: University of Las Palmas de Gran Canaria.
- Bernardini, Silvia (2007): *Collocations in Translated Language: Combining Parallel, Comparable and Reference Corpora*. Paper presented at the fourth Corpus Linguistics conference held at the University of Birmingham, 27–30 July 2007.
- Corpas Pastor, Gloria (1996): *Manual de fraseología española*. Madrid: Gredos.
- Gurrutxaga, Antton (2014): *Idiomatikotasunaren karakterizazio automatikoa: Izena+Aditza konbinazioak*. PhD thesis. University of the Basque Country.
- Häcki-Buhofer, Annelies, Marcel Dräger, Stefanie Meier & Tobias Roth (2014): *Feste Wortverbindungen des Deutschen. Kollokationenwörterbuch für den Alltag*. Tübingen: Francke.
- Hulden, Mans (2009): Foma: A finite-state toolkit and library. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, 29–32.
- Ibarretxe-Antuñano, Iraide (2006): Estudio lexicológico de las onomatopeyas vascas: el euskal onomatopeien hiztegia: euskara-ingelesera-gaztelania. *Fontes Linguae Vasconum: Studia et Documenta* 101, 147–162.
- Ibarretxe-Antuñano, Iraide (2012): Análisis lingüístico de las onomatopeyas vascas (Linguistic analysis of Basque onomatopoeia). *Oihenart: Cuadernos de Lengua y Literatura* 27, 129–177.
- Ibarretxe-Antuñano, Iraide & Alfontso Martínez Lizarduikoa (2006): *Hizkuntzaren bihotzean: euskal onomatopeien hiztegia*. Donostia-San Sebastian: Gaiak.

- Iñurrieta, Uxo, Itziar Aduriz, Arantza Diaz de Ilarraza, Gorka Labaka & Kepa Sarasola (2016): Izen+aditz konbinazioen itzulpenaz eta tratamendu konputazionalaz. *Senez* 47, 237–249.
- Moon, Rosamund (1998): *Fixed Expressions and Idioms in English*. Oxford: Clarendon Press.
- Sanz-Villar, Zuriñe (2015): *Unitate fraseologikoen itzulpena: Alemana euskara. Literatur testuen corpusean oinarritutako analisisa*. PhD thesis. Servicio Editorial de la UPV/EHU.
- Sanz-Villar, Zuriñe (2018): Interference and the translation of phraseological units in a parallel and multilingual corpus. *Meta* 63 (1), 72–93.
- Schuchardt, Hugo (1925): *Das Baskische und die Sprachwissenschaft*. Viena: Hölder-Pichler-Tempsky A.-G.
- Sinclair, John (1987): The nature of the evidence. In John Sinclair (ed.), *Looking Up: An Account of the COBUILD Project in Lexical Computing*, 150–159. London: Collins.
- Toury, Gideon (2012): *Descriptive translation studies and beyond*. Amsterdam, Philadelphia: John Benjamins.
- Uribarri, Ibon (2011): Dialectics of opposition and construction: Translation in the Basque Country. In Dimitris Asimakoulas & Margaret Rogers (eds.), *Translation and Opposition*, 247–264. Bristol: Multilingual Matters.
- Urizar, Ruben (2012): *Euskal lokuzioen tratamendu konputazionala*. PhD thesis. University of the Basque Country.