Georgina Alvarez-Morera, Jordi Ginebra, and Isabel Oltra-Massuet

Modification in *give* light verb constructions

A corpus-based study in Germanic and Romance languages

1 Introduction

Since Wierzbicka (1982), light verb constructions (LVCs) have been regarded as the telic counterpart (1a) of atelic verbs (1b), thus accounting for the existence of such similar constructions. However, that cannot be the whole story, since the number of the nominal element affects the telicity of the construction, as illustrated in the contrasts (1a–(1c)).

(1) a. *Melinda had a thought* [+telic]
b. *I thought that the advice I was giving my clients was special.* [–telic]
c. *She also had thoughts of suicide.* [–telic]
(Bonial and Pollard, 2020, 15)

In more recent approaches, the higher frequency of LVCs tends to be attributed to their more flexible modification properties, at least in English (Leech et al., 2009, 179; Huddleston and Pullum, 2002, 293; Bonial and Pollard, 2020, 18). Yet, corpus-based studies, such as Storrer (2007) or Levin and Ström Herold (2015), have shown that modification is not common in LVCs in Germanic languages (e.g. English, German and Swedish). The fact that previous corpus findings are all based on the same language family hinders the possibility of finding the underlying generalization across languages.

Thus, there is a need for corpus-based studies that conduct a cross-linguistic analysis that includes additional language families with a focus on the nominal component of the LVC (as pointed out in Alvarez-Morera, 2023). With this aim, the present study investigates *give*-LVCs across Germanic (English, *give*, and German, *geben*) and Romance languages (Catalan, *donar*, and Spanish, *dar*). Taking a random sample of 6,794 tokens of *give*-LVCs from online annotated corpora for the four languages within the time frame of the 21st Century as a basis, we review the morphosyntactic properties of the nominal element, with a special focus on their modifi-

Georgina Alvarez-Morera, University of Oxford
Jordi Ginebra, Universitat Rovira i Virgili
Isabel Oltra-Massuet, Universitat Rovira i Virgili

https://doi.org/10.1515/9783111388878-004

cation patterns. After the quantitative analysis of the corpus sample in terms of frequency and modification patterns, we observe that the majority of LVCs in all four languages appear unmodified, although there are significant differences within and across language families: modified nominal elements are most frequent in English, while German patterns with Romance languages in that it shows lower percentages of modification of about a third of the instances. Interestingly, Wiskandt (this volume) also concludes that no significant contrast can be established between Germanic LVCs and their Romance counterparts.

We take these findings as a basis for an analysis of LVCs within a neo-constructionist approach to argument structure (Acedo Matellán, 2010, 2016, based on Hale and Keyser, 53–109, 2002), which focuses on the determination and modification properties of the nominal in LVCs. The results point towards a compositional hierarchy of the nominal domain based on Grimshaw's (2005) notion of extended projection: the more independent the nominal from the light verb (LV), the more structure is projected within the nominal domain. Unlike previous studies which focus on the role of the light verb to explain restrictions on the nominal element (Mendívil, 1999), our claim is that it is the nature of the syntactic nominal phrase that establishes the combinatorial patterns of the LVC as a whole.

The structure of the chapter is as follows. In section 2, we briefly review the analysis of LVCs, focusing on how modification in LVCs has been dealt with in previous studies. In section 3, we describe the methodology employed in our corpus-based study on Germanic and Romance languages. Section 4 presents the quantitative results of the study, as well as the statistical analysis 4.2, and discusses the results 4.3. Taking our corpus-based, empirical results as a basis, section 5 establishes a typology of LVCs based on the presence of determination and modification. It further outlines a syntactic account of the various LVCs that is grounded on the layered approach to the nominal domain within a neo-constructionist approach to argument structure. Section 6 concludes and suggests directions for future research.

2 Properties of LVCs across frameworks

This section provides a brief overview of how LVCs are typically defined, their basic properties and what the consensus is on the presence of nominal modification.

The term *light verb* was first coined by Jespersen (1942, 117) to refer to verbo-nominal constructions in English like *have a sigh* or *take a shower*, where a deverbal noun is combined with a semantically impoverished verb to form a complex predicate. After this initial definition, different proposals within various frameworks have examined these structures. While in the 1970s and 1980s, there was certain

consensus in considering the LVs completely devoid of predicational force and the argument structure of LVC as inherited from the nominal element (Cattell, 1984; Grimshaw and Mester, 1988), more recent accounts suggest that the LV also contributes to the argument and event structure of the construction (Ginebra, 2003; Ramchand, 2014; Acedo Matellán and Pineda, 2019).

It is also important to bear in mind that LVCs usually have a synthetic verb which a priori conveys the same meaning (Sanromán Vilas, 2009; Butt, 2010): *walk* for *take a walk* or *pasear* for *dar un paseo* Sp.'take a walk'. However, this equivalence is not always available within a particular language and there are four possibilities (Piera and Varela, 1999): (i) the synthetic counterpart is equivalent in both its morphology and semantics (2), (ii) there is morphological affinity but no semantic equivalence (3), (iii) there is only a semantic equivalence (4), and (iv) there is no synthetic counterpart (5).[1]

(2) a. Eng. *take a shower – to shower*
b. Ger. *Antwort geben* lit. 'give answer' – *antworten* 'answer'
c. Sp. *dar besos* lit. 'give kisses' – *besar* 'kiss'
d. Cat. *fer una còpia* lit. 'make a copy' – *copiar* 'copy'

(3) a. Eng. *give voice – to voice* 'express'
b. Sp. *hacer fiesta* lit. 'do holiday' – *festejar* 'celebrate'
c. Cat. *donar propina* lit. 'give tip' – *propinar* 'give/administer'

(4) a. Eng. *make a mistake – to err*
b. Ger. *Platz nehmen* lit. 'take (a) place' – *setzen* 'put'
c. Sp. *hacer punto* lit. 'do stitch' – *tricotar* 'knit'
d. Cat. *donar un mastegot* lit. 'give a slap' – *pegar* 'hit'

(5) a. Eng. *take advantage*
b. Ger. *Abstand nehmen* 'reject' (lit. take distance)
c. Sp. *hacer deporte* 'do sport' (lit. do sport)
d. Cat. *donar conferéncia* 'lecture' (lit. give conference)

Even when there is a synthetic verb, as in (2–4), the semantic equivalence between the pairs is not always complete in every context. As mentioned above, Wierzbicka

1 German LVCs with the nominal introduced by a preposition do not generally have a corresponding paraphrasis available, as pointed out in Fleischhauer (2021, 2022) and, hence, they fall into the (iv) pattern. These PP LVCs have been widely studied in German because they appear to be the most representative type of LVCs in that language; however, they are less present in the rest of the languages included in this study (English, Catalan and Spanish). Therefore, the focus of our work focuses on the pattern of a LV followed by a noun phrase.

(1982, 764) points to aspectual differences between English LVCs and their synthetic counterpart as the key motivation for the use of LVCs.

(6) a. *I had a look at all the files in ten minutes.*
b. ?*I looked at all the files in ten minutes.*

(Wierzbicka, 1982, 764)

English LVCs have been further investigated by Bonial and Pollard (2020) on the basis of corpora and they find telicity differences only when the nominal element is singular (7b). The instance in (7c) shows that a plural noun does not express a telic event.

(7) a. *I thought that the advice I was giving my clients was special.* [–telic]
b. *Melinda had a thought: Maybe there would be some way for her husband to collect Mann's DNA* [...] [+telic]
c. *She also had thoughts of suicide.* [–telic]

(Bonial and Pollard, 2020, 15)

Other LVs in English are combined with nouns that are already inherently telic, as in (8) and (9), so the choice of a LVC over the synthetic verb cannot be said to have an aspectual motivation.

(8) a. *Twelve hours into the siege, darkness was beginning to set in, and police decided to make a move.* [+telic]
b. *The governor made a decision to release his tax records.* [+telic]

(9) a. *She appeared with me on VH1 'Celebrity Rehab.'* [+telic]
b. *Bahrain's King Hamad made a rare appearance on television.* [+telic]

(Bonial and Pollard, 2020, 16)

Therefore, there is systematic relation between the aspect of the event denoted by the noun, and the telicity of the whole construction. Yet, as pointed out in Storrer (2007, 186), the choice between the LVC and the corresponding synthetic verb "is not as arbitrary as it has been assumed in the literature", hence a thorough description of which factors favors one or the other type of construction is necessary.

When analysing the paraphrases of synthetic verbs with LVCs in German, Storrer (2007, 180) finds that less than 50% of the instances occurred in contexts compatible with the LVC. LVCs seem, thus, to be semantically more specific, since they can only paraphrase the synthetic verb in those instances where they share the specific meaning.

These observations have also been made for other languages, such as Italian (Bratánková, 2013), where some LVCs have developed a lexicalized meaning inde-

pendent from the synthetic verb: *fare festa* 'make party' vs. *festeggiare* 'celebrate' or *dare peso* 'give importance' (lit. give weight) vs. *pesare* 'weigh'. This does not apply to all LVCs, since some LVCs are still used as an analytic variant of the synthetic verb (*fare una telefonata* 'make a call' vs. *telefonare* 'call'). However, the use of the LVC in Italian is prevalent where there is a qualitative denotation of the action through the modification of the nominal. This finding is in line with the results presented by Bonial and Pollard (2020, 19–21) for English LVCs.

Nonetheless, Storrer's (2007, 184) corpus study on German LVCs finds that a considerable number of nouns are modified by adjectives to different degrees:

Tab. 1: Noun modified by adjectives (Storrer, 2007, 184).

LVC	Instances	Modified by an adjective
Hilfe leisten 'provide assistence'	310	85 (27.4%)
Unterricht erteilen 'give lecture'	122	34 (27.9%)
Wirkung ausüben 'exert influence'	275	196 (71.3%)
Absage erteilen 'give rejection'	82	42 (51.2%)

Some cases of adjectival modification in LVCs cannot keep their original meaning when transformed into an adverbial modifier of the synthetic counterpart, as in (10), where the modifier is ambiguous or does not sound natural. Also, there are cases, as in (11), where the adjective does not have an adverbial equivalent.

(10) a. *We had – we had a really good laugh and then things happened.*
b. ?*We laughed really well and then things happened.*

(11) a. *President Bush gave a very big speech in September of that year – transferring those detainees to Guantanamo.*
b. **President Bush spoke very bigly in September of that year – transferring those detainees to Guantanamo.*

(Bonial and Pollard, 2020, 20)

The ease of modification of the noun in contrast with the adverbial modification of the synthetic verb has been considered one of the reasons for the high productivity of LVCs cross-linguistically (Huddleston and Pullum, 2002, 293, Rácz et al., 2014, 5–6, Bonial and Pollard, 2020, 21).

Therefore, Levin and Ström Herold (2015) examine the modification of LVCs in Germanic languages (English, German and Swedish), and they find that the most frequent modification of the nominals in LVCs is adjectival modification. However, this is not present in the majority of instances (see Table 2): the LV with the highest

percentage is *give* (32% in English, 19% in Swedish and 24% in German), while *take*-LVCs are comparatively less modified (22% in English, 13% in German and only 5% in Swedish).

Tab. 2: LVCs with adjectival modification (Levin and Ström Herold, 2015, 20).

	English	German	Swedish
give / geben / ge	32% (58/184)	24% (23/94)	29% (41/140)
take / nehmen / ta	22% (39/174)	13% (9/71)	5% (13/265)

Thus, one of the aims of our study is to contribute new and comparable data from a larger set of languages that examines and contrasts the properties of modification in Germanic (English, German) and Romance (Catalan and Spanish) languages.

3 Data and methodology

The present study is based on data from online available corpora in the four languages under study: *Corpus of Contemporary American English* (COCA), *Das Digitale Wörterbuch der Deutschen Sprache* (DWDS), *Corpus textual informatitzat de la llengua catalana* (CTILC), and *Corpus del Español del Siglo XXI* (CORPES XXI). Although there are certain differences between these corpora, especially based on the number of words and the time span they cover (see Table 3), the time frame set for the extraction was the 21st Century in order to concentrate on the most contemporary data of the four corpora and maximally avoid potential biases.[2]

Among the various LVs available, the verb *give* has exhibited stability in terms of frequency since early modern English, as indicated by Butt (2010, 68) and Elenbaas (2013, 66). In contrast, the German verb *geben* does not hold the status of being the most prevalent LV (Bruker, 2011, 45). However, some researchers (i.e. Levin and Ström Herold, 2015), have opted for *geben* over its formal counterpart *erteilen* when comparing it with the English LV *give*. Turning to Romance languages, the LV *dar* in

2 There are no corpora in these four languages that are the same size and cover the same time span and the same textual typology, so we have considered those that are the most homogeneous despite certain divergences and have constrained the time span. In all cases, there is some institution behind these corpora that is responsible for the text selection, annotation and the open access availability of their data (the scholar Mark Davies for COCA, the Berlin-Brandenburg Academy of Sciences and Humanities for DWDS, the Institute of Catalan Studies for CTILC, and the Royal of the Academy Spanish Language for CORPES XXI).

Tab. 3: Comparison of the corpora.

Lang.	Corpus	Size	Genres	Time
English	COCA	+1 billion words of text (25+ million words each year 1990-2019)	8 genres: spoken, fiction, magazines, newspapers, academic texts, blogs, and TV and movies subtitles	2000–2019
German	DWDS	28 billion words of text (1900-2020)	6 genres: spoken, fiction, newspapers, academic texts, blogs, and TV and movies subtitles	2000–2020
Catalan	CTILC	+100 million words of text (1832-2018)	3 genres: fiction, newspapers and academic texts, but there are no instances of spoken language	2000–2018
Spanish	CORPES XXI	312 million words of texts (21st Century)	6 genres: spoken, fiction, newspapers, academic texts, blogs and others	2000–2020

Spanish and *donar* in Catalan rank among the most frequently used verbs. In Spanish, *dar* holds the distinction of being the most widespread verb in the language, according to Sánchez Rufat (2016, 118), and it also represents the LV with the highest number of potential combinations, as noted by De Miguel (2011, 142). Conversely, in Catalan, *donar* is not the most popular LV, it is *fer* 'make/do', but *donar* still maintains a notable frequency, as highlighted in Ginebra and Navarro (2015, 222).

The sample was obtained by searching for the collocation of the LV *give* (and its corresponding translation in each language: *geben*, *donar* and *dar*) and the specific nominal element with which it builds an LVC. The final selection of *give*-LVCs to be analyzed for each language was set to a maximum of 15 LVCs per language, with a range from highly frequent LVCs to less frequent. The highest number of lemmas (n=15) is found in Spanish *dar* LVCs, whereas the number of Catalan (n=14), English (n=13) and German (n=9) LVCs is slightly lower (Table 4). This is due to the fact that LVCs are more diverse in Spanish than in other languages, while in German LVCs with *geben* followed by a noun phrase is not a common pattern within the language (Rácz et al., 2014, 5).

After manually excluding instances of non-LV structures, for every LVC a sample of up to 200 occurrences was analyzed. The LVCs were organized according to the degree of frequency in the corpora under the label frequency: high, medium,

Tab. 4: LVCs analysed in the four languages under study.

	High frequency	Medium frequency	Low frequency
give (English)	*advice, speech, answer, example*	*try, hug, smile, kiss*	*notice, nod, push, chase, sigh*
geben (German)	*Hinweis* 'hint', *Antwort* 'answer'	*Rat* 'advice', *Unterricht* 'lesson', *Erlaubnis* 'permission'	*Kuss* 'kiss', *Warnung* 'warning', *Beschreibung* 'description', *Versprechen* 'promise'
donar (Catalan)	*suport* 'support', *cop* 'blow', *resposta* 'answer', *volta* 'stroll', *explicació* 'explanation'	*consell* 'advice', *permís* 'permission'	*empenta* 'push', *definició* 'definition', *conferència* 'conference', *ajuda* 'help', *bufetada* 'slap', *pallissa* 'beating', *bes/ada* 'kiss'
dar (Spanish)	*vuelta* 'stroll', *paso* 'step', *respuesta* 'answer', *beso* 'kiss'	*golpe* 'blow', *clase* 'lesson', *salto* 'jump', *cambio* 'change', *consejo* 'advice'	*instrucción* 'instruction', *giro* 'turn', *permiso* 'permission', *ayuda* 'help', *bofetada* 'slap', *autorización* 'authorization'

and low.[3] Also, all examples were classified according to the following grammatical factors:[4]

(i) determination (yes/no),
(ii) modification (different classes: unmodified (UNMOD), adjectival (Adj), prepositional phrase (PP), relative clauses (Rel), noun classifiers (NC), genitive (Gen), or various, that is, a combination of two of the former),
(iii) number of the noun (singular/plural).

The final database of *give*-LVCs instances consists of 6,789 tokens of LVC with some differences in the distribution among the four languages (Table 5); although these divergences are consistent with the differences presented in the size of the corpora. The variables were coded in all examples, and then fed into the Lancaster Stats Tools online.[5] First, chi-square tests were used as an initial means to identify potentially

3 The degree of frequency of the LVC has been decided on the co-occurrence between the LV and the different nouns according to the search tool for collocates in each corpus. The list was divided into quartiles and the LVCs were selected randomly from each quartile (high, medium, and low).

4 Some methodological limitations must be pointed out. The manual coding for variables is a usual method in corpus linguistics but it might lead to measurement errors (Egbert et al., 2020). The possible errors are compensated with a high number of instances. All remaining errors are the researchers' own responsibility.

5 Available at: http://corpora.lancs.ac.uk/stats/toolbox.php, last access 05/12/2024.

significant from non-significant variables. Second, a logistic regression analysis was performed to find out which features were the strongest predictors of modification. We report the results in the next section.

Tab. 5: Distribution of the number of LVCs of the database for each language.

	English	German	Catalan	Spanish	TOTAL LV
give / geben / donar / dar	2,030	995	931	2,842	6,798

4 General results

4.1 Overview

The analysis of the data on modification in *give*-LVCs shows that LVCs with modified nominal elements are most frequent in English (53.26%), while German (31.76%) and the Romance languages show slightly lower percentages of modification: 28.57% in Catalan and 27.69% in Spanish (Figure 1). The majority of *give*-LVCs appear unmodified in these three languages. This distribution is statistically significant when all languages are compared ($\chi^2(3)= 376.95$, $p<.0001$, Cramer's $V=0.236$, medium effect), but not within both language families: when Germanic languages are contrasted, statistical tests prove a significant difference ($\chi^2(1)=124.09$, $p<.0001$, Cramer's $V=0.203$, small effect). However, the difference is not significant in the case of the two Romance languages ($\chi^2(1)=0.27$, $p=0.60$, Cramer's $V=0.008$, negligible effect).

The results thus point towards a difference in the tendency to appear modified between English and the other three languages, as shown in Table 6. The modification subclasses tagged are: no modification (Unmod), adjectival (Adj), prepositional phrases (PP), relative clauses (Rel), noun classifiers (NC), genitive (Gen), and various. As seen in Table 6, adjectival modification is the most common in all languages: 41.17% in English (versus 12.09% of the rest of modification subclasses), 27.74% in German (versus 4.02% of the rest of modification subclasses), 19.33% in Catalan (versus 9.24% of the rest of modification subclasses), and 19.56% in Spanish (versus 8.13% of the rest of modification subclasses). In fact, relative clauses (Rel), noun classifiers (NC), genitive (Gen) and a combination of subclasses (Various) are residual in the four languages (below 3% of instances in all languages), where NC and Gen are only found in Germanic languages. However, prepositional phrases (PP)

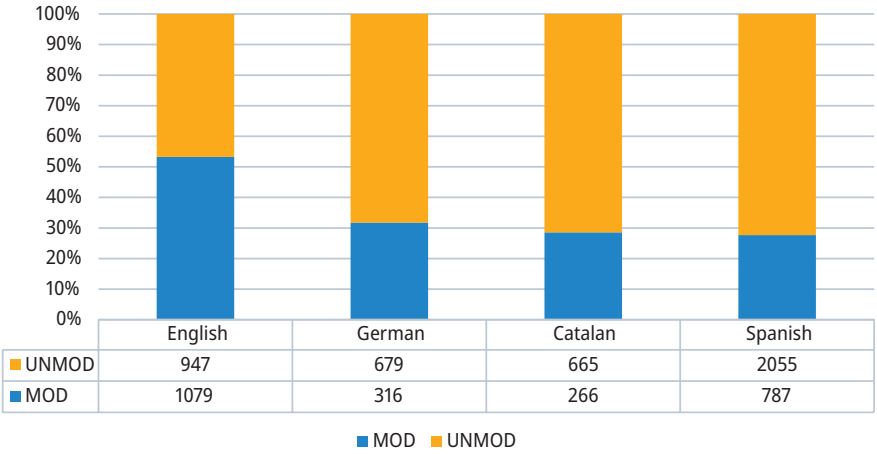


Fig. 1: Distribution of modification in *give*-LVCs.

present similar percentages in the Romance languages, as exemplified in (12): 6.55% in Catalan *donar*-LVCs (12a), and 6.86% in Spanish *dar*-LVCs (12b). These differences in the choice of Germanic languages for NC and Gen and Romance languages for PP reflect the general nominal modification patterns of these language families.

(12) a. *i comença a donar-li cops de puny.*
and starts to give-3SG.DAT hits of fist
'and starts punching him' (Catalan, CTILC, 2001)

b. *al dar instrucciones de neutralidad a la Guardia Civil.*
at.the.SG give instructions of neutrality to the Guard Civil
'when giving neutrality instructions to the Police'
(Spanish, CORPES XXI, 2001)

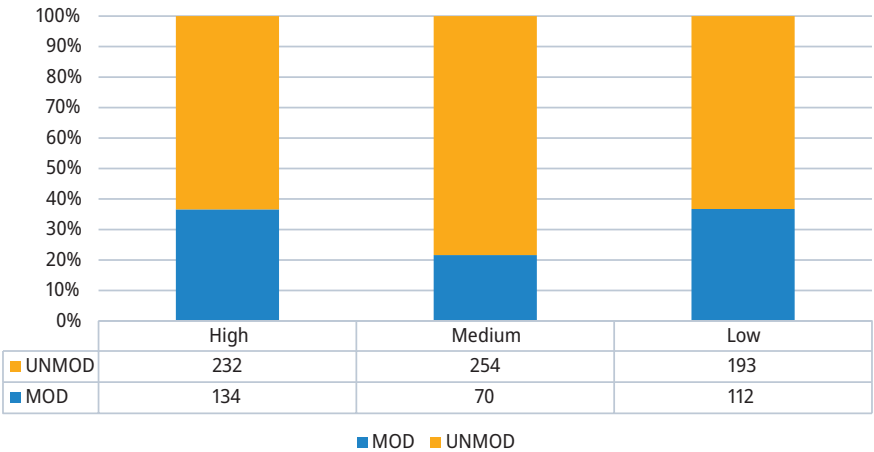
Tab. 6: Distribution of modification subclasses in *give*-LVCs (all four languages)

Modification	English	German	Catalan	Spanish
Unmod (No)	947 (46.74%)	679 (68.4%)	665 (71.42%)	2055 (72.31%)
Adj	834 (41.17%)	276 (27.74%)	180 (19.33%)	556 (19.56%)
PP	61 (3.01%)	3 (0.30%)	61 (6.55%)	195 (6.86%)
Rel	36 (1.78%)	13 (1.30%)	12 (1.29%)	24 (0.84%)
NC	59 (2.91%)	1 (0.10%)	-	-
Gen	29 (1.43%)	11 (1.11%)	-	-
Various	60 (2.96%)	12 (1.21%)	13 (1.40%)	12 (0.43%)
TOTAL	2026	995	931	2842

It can be concluded that general differences in the modification of the noun in *give*-LVCs are statistically validated.

Furthermore, a closer look at the correlation between the degree of frequency of LVCs in the corpora and the presence or lack of modification is needed to determine whether frequency has an influence on the overall results.[6]

Figure 3 shows that English *give*-LVCs with high frequency appear modified in most cases (61.42%); those with medium frequency are modified in half of the cases (51.43%); low-frequency LVCs are modified in only 44.10% of the instances. Interestingly, the differences between the degrees of frequency are statistically significant ($\chi^2(2)=40.23$, $p<.0001$, Cramer's $V=0.141$, small effect).



	High	Medium	Low
UNMOD	232	254	193
MOD	134	70	112

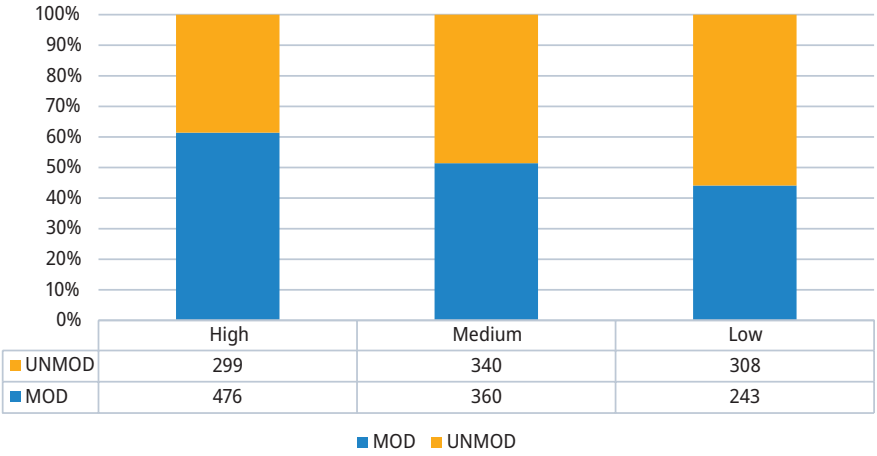
Fig. 2: Distribution of modification in German *geben*-LVCs according to the degree of frequency (high, medium or low).

Hence, the results for English point to a correlation between degree of frequency and levels of modification: the more frequent an LVC, the more likely it will appear modified.

In a similar line, Figure 2 shows that the German high-frequent *geben*-LVCs are modified as often as the low frequency group (36.61% and 36.72% respectively), and to a lesser extent the medium frequent (21.60%). The differences between the de-

6 Frequency is treated here as a categorical variable although it is not a naturally categorial variable for an intention of homogeneity between the four corpora. Moreover, token frequencies are the most widely used corpus statistic, but they are problematic and corpus-dependent (as pointed out in Gries, 2023, 82).

grees of frequency are, again, statistically significant ($\chi^2(2)=22.86$, p<.0001, Cramer's V=0.152, small effect).



	High	Medium	Low
UNMOD	299	340	308
MOD	476	360	243

Fig. 3: Distribution of modification in English *give*-LVCs according to the degree of frequency (high, medium or low).

Turning to Romance languages, Figure 4 shows that the distribution of modification in Catalan *donar*-LVCs presents a different trend that radically contrasts with Germanic languages. The high-frequency Catalan *donar*-LVCs are modified in only 28.12% of instances and medium-frequency in 20.15%, which contrast with the low-frequency LVCs that appear modified in 40% of instances.[7] The difference between the groups of frequency is statistically significant ($\chi^2(2)=12.72$, p=0.001, Cramer's V= 0.117, small effect).

A closer analysis of the morphosyntactic properties of Catalan *donar*-LVCs shows that there are certain restrictions to adjectival modification which are related to the bareness of the nominal element (Espinal and McNally, 2011, 113). Bare nominals tend to be modified by relational adjectives (13a), and they only accept qualitative adjectives in restricted contexts (13a).

7 The percentual divergence can be explained by the lower number of examples extracted for the low-frequency LVCs, which is due to the examples available in the CTILC corpus for the 21st Century.

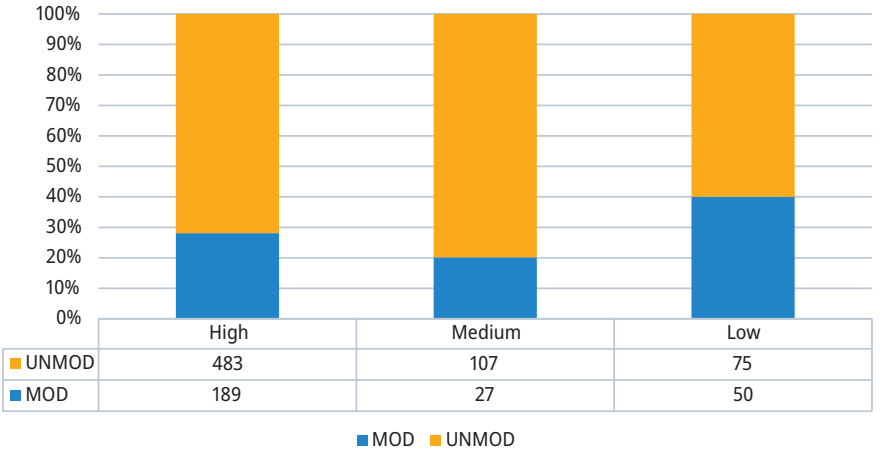


Fig. 4: Distribution of modification in Catalan *donar*-LVCs according to the degree of frequency (high, medium or low).

(13) a. *i dóna suport tècnic i personal a la resta d'unitats.*
and gives support technical and personal to the rest of-units
'and [it] gives technical and personal support to the rest of the units.'
(Catalan, CTILC, 2010)

b. *catalans i balears es donaran suport mutu.*
Catalan.PL and Balearic.PL SE give.3PL.FUT support mutual
'Catalan and Balearic [people] will give mutual support'
(Catalan, CTILC, 2005)

These restrictions could have an impact on the results of Catalan *donar*-LVCs which accept the nominal to appear bare. In the sample of Catalan *donar*-LVCs, high and medium-frequency LVCs include constructions which accept both determined and bare nominals (14a), as well as others which only accept the nouns with a determiner in singular (14b); while low-frequency LVCs mainly accept determined nominals (15b) and only two instances can combine with bare nominals (15a).[8]

8 The fact that determined nominals show fewer restrictions in accepting modification might have influenced the results and the differences presented in Figure 4, where the low-frequency *donar*-LVCs show the highest percentages of modification.

(14) a. *donar* (*un*) *suport* 'give (a) support', *donar* (*una*) *resposta* 'give (an) answer', *donar* (*un*) *consell* 'give (an) advice', *donar* (*un*) *permís* 'give (a) permission'

b. *donar* *(*un*) *cop* 'give *(a) blow', *donar* *(*una*) *explicació* 'give *(an) explanation', *donar* *(*una*) *volta* 'give (a) walk'

(15) a. *donar* (*una*) *ajuda* 'give (a) help', *donar* (*una*) *conferència* 'give (a) conference'

b. *donar* *(*una*) *empenta* 'give *(a) push', *donar* *(*una*) *definició* 'give *(a) definition', *donar* *(*una*) *bufetada* 'give *(a) slap', *donar* *(*una*) *pallissa* 'give *(a) beating', *donar* *(*un*) *bes/besada* 'give *(a) kiss'

For Spanish *dar*-LVCs, Figure 5 shows that medium-frequency LVCs have the highest percentage of modified instances (36.16%), followed by the high-frequency LVCs (25.12%), and low-frequency are the least modified (23.27%). The difference between the frequency groups is, again, statistically significant ($\chi^2(2)=45.02$, $p<.0001$, Cramer's V = 0.126, small effect).

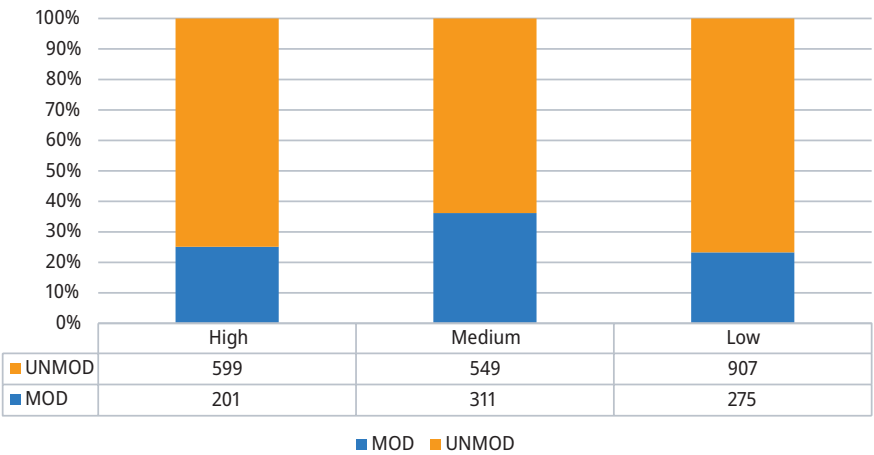


Fig. 5: Distribution of modification in Spanish *dar*-LVCs according to the degree of frequency (high, medium or low).

In sum, the degree of frequency of LVCs in corpora has proven to have contradictory effects on the results. There is a positive correspondence between high-frequency LVCs and a higher presence of modification in the languages with bigger corpora, English and Spanish, although the proportion is lower in Spanish. This tendency is also relevant in German LVCs but to a lesser extent since low-frequency *geben*-LVCs

are also equally modified. In contrast, there is a negative correspondence between frequency of LVCs and their possibilities to be modified in Catalan *donar*-LVCs, as it is the low-frequency constructions which are the most frequently modified.

4.2 Statistical analysis

In order to confirm the corpus-based, empirical trends attested in the previous section, we employed a binomial logistic regression as statistical method, which is performed with the most significant variables, to determine which of them are the strongest predictors for modification in LVCs. The implementation of this statistical test aims at examining the extent to which the morphosyntactic variables considered in the previous sections (number and determination), as well as the degree of frequency, can predict the patterns of modification attested.[9]

The result of the logistic regression analysis is given in Table 7, which was overall statistically significant (LL: 442.57; $p < 0.0001$).

Tab. 7: Output of the logistic regression model for predictor factors in modification with *give* LVCs. Logistic Regression Model: Response variable: Mod (unmodified |modified) (4344 |2448). Overall model statistics: Likelihood ratio test (LL): 442.57 ($p < 0.0001$); C-index: 0.65; Nagelkerke R2: 0.09; AIC: 8454.56.

Coefficients	Estimate (log odds)	Standard Error	Z value (Wald)	p-value
(Intercept)	0.299	0.058	5.110	0.000
Langb_de	-0.879	0.082	-10.757	0.000
Langc_ca	-0.978	0.089	-10.960	0.000
Langd_es	-0.998	0.063	-15.850	0.000
Detb_no	-0.461	0.065	-7.058	0.000
Numb_plur	0.086	0.069	1.239	0.215
Freqb_medium	-0.078	0.065	-1.206	0.228
Freqc_low	-0.258	0.066	-3.893	0.000

According to the regression analysis of the corpus material tabulated in Table 7, the variables language and determination are the strongest distinguishing factors. First, three languages, German, Catalan and Spanish, are all less likely than English to be

9 Predictors included in the model are language (Lang), determination (Det), number (Num), and frequency (Freq). No predictor interactions are included. For the calculation, the Lancaster Stats Tools online (Brezina, 2018) is used, which includes the possibility to generate a logistic regression model in the section 'Lexico-grammar'. Available at: http://corpora.lancs.ac.uk/stats/toolbox.php

modified. Second, bare nouns are also less likely than nouns introduced by a determiner to be modified. In contrast, number does not show any significant predicting power.

Regarding the predictor variable frequency, only low frequent LVCs show a negative estimate, which indicates that it is less likely to be modified than the reference variable (highly frequent LVCs). The comparison with medium-frequency LVCs does not show significant results.

Thus, the results of the statistical model are fully in line with the findings reported in the descriptive presentation of the data in section 3.

4.3 Interim discussion

The main purpose of this corpus study was to review the morphosyntactic properties of the nominal element in LVCs, with a special focus on their modification patterns. The results show that modification is clearly more frequent in English, while it is less frequent in German, Catalan, and Spanish to different degrees. Moreover, the corpus data show very little prominence of modification subclasses other than adjectival modification in the four languages. This is in line with previous research by Levin and Ström Herold (2015, 24) who discard the in-depth analysis of clausal modifiers because they represent only 5% of instances with *give*-LVCs and 2% in *geben*-LVCs. The proportions are even lower in our data: in English *give*-LVCs, relative clauses represent 41 out of 2026 instances (2.02%); in German *geben*-LVCs, 14 out of 995 instances (1.41%). In the case of Romance languages, relative clauses also show residual presence: in Catalan *donar*-LVCs, 12 out of 931 (1.21%), and in Spanish *dar*-LVCs, 25 out of 2842 (0.88%).

Regarding PP modification, however, there are no previous studies which can be compared to, because the focus on this kind of modification is not relevant for Germanic languages. In fact, nominal modification through PPs is low to residual in Germanic languages: 5.53% in English *give*-LVCs (112 out of 2026) and only 0.60% in German *geben*-LVCs (6 out of 995). In contrast, PP modification shows a higher presence in the two Romance languages under study: 7.95% in Catalan *donar*-LVCs (74 out of 931) and 10.12% in Spanish *dar*-LVCs (205 out of 2026). Even if PP-modification is more present in the two Romance languages studied, it is still low and only more relevant than adjectival modification with certain nouns: *cop* 'blow', *conferència* 'conference' and *permís* 'permission' in Catalan; *clase* 'lesson' and *permiso* 'permission' in Spanish.

The results on adjectival modification in *give*-LVCs in Germanic languages show that English LVCs are the most frequently modified in 834 out of 2026 instances (41.16%), and German presents lower numbers in 276 out of 995 instances (27.73%).

In Levin and Ström Herold (2015, 20), the results are slightly lower: 32% in *give*-LVCs in English (58/184) and 24% in *geben*-LVCs in German (23/94). However, their sample was smaller and less representative, since it was restricted to the genre of fiction. At the same time, the levels of modification found in our study are slightly lower than the findings by Bonial and Pollard (2020, 16–17) which represented more than 64% on average. However, their results are not comparable to ours either, due to the kind of modifiers included in their analysis, which involved quantifiers, determiners, and relative clauses alongside adjectival modification.

To our knowledge, all previous studies have only included data from Germanic languages, so the results for Catalan *donar*-LVCs and Spanish *dar*-LVCs are novel. In this case, results on adjectival modification are more similar within the Romance language family, but they sharply contrast with the Germanic languages in that they represent a much lower percentage: 180 out of 931 instances in Catalan (19.33%) and 556 out of 2,842 in Spanish (19.56%).

Regarding the degree of frequency of the LVC lemmas, this parameter has contradictory effects on the results, as clearly shown by the logistic regression analysis. Hence, the degree of frequency is a parameter which cannot homogeneously predict the levels of modification in LVCs cross-linguistically without considering the properties of particular LVCs. In fact, previous studies on the use of LVCs versus their synthetic counterparts (i.e. Storrer, 2007; Sanromán Vilas, 2009; Bratánková, 2013) point towards the tendency of frequent LVCs to develop a lexicalized meaning, which would favor their use over that of a synthetic verb. In our data, however, a majority of specific LVCs (regardless of their degree of frequency) combine with a wide range of adjectives, both relational and qualitative, and the tendencies of co-occurring adjectives and nouns are residual.

5 A typology of LVCs across Germanic (English/ German) and Romance (Catalan/ Spanish)

Based on the restrictions and interactions found between modification – relational or qualitative – and determination detected in our corpus-based study, we propose a corpus-based typology of LVCs across Germanic (English/German) and Romance (Catalan/Spanish) languages based on the morphosyntactic structure of these constructions taking into account the deverbal nature of the nominal component, as well as additional restrictions on adjacency and the presence of quantifiers, as summarized in Table 8.

In this section, we discuss the properties of each type and subtype of LVC and suggest a syntactic analysis that focuses on the compositional hierarchy within the

Tab. 8: Typology of LVCs.

	Type 1	Type 2		Type 3
		Deverbal	Non-deverbal	
Deverbal	Yes	Yes	No	—
Determination	Yes	No	No	No
Quantification	Yes	Yes	Yes	No
Adjectival modification	Yes	Yes	Yes (relational)	No
Adjacency	No	No	No	Yes
Examples	*dar un abrazo efusivo* (Sp., give a warm hug)	*dar respuesta afirmativa* (Sp., give positive answer)	*donar consell econòmic* (Cat., give economic advice)	*dar vuelta* el partido (Sp., turn the match around)

DP, where the structure above N determines the degree of integration between the nominal and the verb. The basic idea underlying our syntactic analysis builds on Grimshaw's (1990; 2005), notion of extended projection. Specifically, we suggest that the more independent the nominal from the light verb, the more structure is projected within the nominal domain.

Unlike previous studies which focus on the role of the light verb to explain restrictions on the nominal element (Mendívil, 1999), our claim is, thus, that it is the nature of the syntactic nominal phrase that establishes the combinatorial patterns of the LVC as a whole. The proposal is further framed within the Hale & Keyser's model of argument structure as developed in Acedo Matellán (2016) and Myler (2016).

5.1 Type 1: LVCs with a determiner and flexible modification

The first type of LVCs that we propose are LVCs with a determiner or a quantifier introducing the noun, in which case modification is flexible; if the noun is deverbal, and modification is adjectival, it may be multiple (16).

(16) a. *I will do my best to give brief and unbiased advice.*
(English, COCA, 2011)

b. *Le dio un largo beso en la boca a su marido*
him gave a long kiss in the mouth to her husband
'She kissed her husband in the mouth for a long time'
(Spanish, CORPES XXI, 2001)

c. *En català, el diccionari Fabra dóna una definició senzilla, però de*
in Catalan the dictionary Fabra gives a definition simple but of
caràcter més enciclopèdic que no pas lingüístic
character more encyclopedic than not not linguistic
'In Catalan, the Fabra dictionary gives a simple definition, but with a more encyclopedic approach than linguistic'
(Catalan, CTILC, 2005)

d. *Nur einmal wollte ihr ein Mann einen sogenannten guten Rat*
only once wanted her a man a so-called good advice
geben
give
'Only once a man wanted to give her a so-called good piece of advice'
(German, DWDS, 2011)

In the case of non-deverbal nouns, adjectival modification is facilitated by the presence of the determiner in the two Romance languages,[10] as exemplified in (17). That is, modification requires the presence of a determiner, in which case there are no restrictions (as noticed by Alonso Ramos, 2004, 198); otherwise, when the nominal is bare, modification is not possible (18).

(17) a. *los senadores dieron el permiso correspondiente para que el*
the senators gave the permission corresponding for that the
gobernante mexicano se ausente del territorio nacional
governor Mexican SE leaves of.the territory national
'the senators gave the corresponding permission for the Mexican ruler to leave the country'
(Spanish, CORPES XXI, 2001)

b. *o bé donava un consell deliberadament ingenu.*
or well gave a advice deliberately naïve
'or he gave a deliberately naïve advice'
(Catalan, CTILC, 2012)

(18) a. *los senadores dieron *(el) permiso correspondiente*
the senators gave the permission corresponding
'the senators gave (the) corresponding permission' [Spanish]

b. *o bé donava *(un) consell ingenu*
or well gave a advice naïve
'or he gave (a) naïve advice' [Catalan]

10 This restriction on bareness is, however, only found in the corpus data of the two Romance languages under study, as will be further developed in the upcoming section 5.2.

Besides the ease in modification, the nouns introduced by a determiner show argumental properties and pass tests, such as the possibility to passivize and pronominalize, exemplified in (19) and ((20) respectively, which have been taken as proof for the referential properties of arguments (Van Valin, 2001, 90, Carnie, 2002, 35–36).

(19) a. *Otros pasos han sido dados con la destitución, averiguación y*
other steps have been given with the dismissal inquiry and
hasta detención de varios funcionarios
until arrest of several civil servants
'Other steps have been taken with the dismissal, inquiry and even arrest of several officials'
(Spanish, CORPES XXI, 2013)

b. *La bufetada havia estat donada i assumida.*
the slap had been given and assumed
'The slap was given and accepted'
(Catalan, CTILC, 2003)

c. *A few examples were given*
(English, COCA, 2012)

d. *Es ist zu vermuten, dass damit die Erlaubnis gegeben ist*
it is to assume that this.with the permission given is
'It is to assume that with this the permission is given'
(German, DWDS, 2012)

(20) a. *Había dado una respuesta y la defendió hasta el final.*
had given a answer and it defended until the end
'He had given an answer and defend it to the end' [Spanish]

b. *El comte va donar el permís a l'abat, però el va donar a contracor.*
the count goes.AUX.PST give the permission to the_abbot but it goes.AUX.PST give at reluctantly
'The Count gave permission to the abbot, but he gave it reluctantly.'
[Catalan]

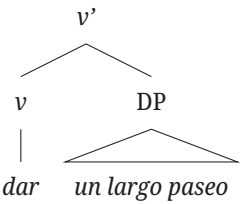
c. *She gave a speech and they remembered it.* [English]

d. *Hat Van der Vaart die Antwort gegeben? Er hat sie endlich gegeben.*
has Van der Vaart the answer given he has it finally given
'Has Van der Vaart given the answer? He has finally given it.'
[German]

Hence, our proposal of analysis for Type 1 LVCs in (21) is that the complement of the LV is a canonical DP within a basic dyadic or transitive construction that would

parallel Hale & Keyser's basic V+N structure for unergative verbs,[11] as developed in Acedo Matellán (2016); Myler (2016), *i.a.*

(21) *dar un largo paseo*
give a long walk
'take a long walk'



This would also be further in line with Bruening's (2015, 58–59) hypothesis that LVCs behave as traditional verb-noun constructions.

5.2 LVCs with bare nominals

Bare nominals in LVCs cannot behave as canonical arguments due to the lack of the determiner (as predicted in Longobardi, 2005, 36). That is, they cannot be passivized (22b) and pronominalization is restricted to the partitive in languages where this pronoun is in use (i.e. Catalan, 22b). Therefore, they cannot be said to take a DP as a complement, unlike Type 1, but they must rather take some other projection below D, as proposed in e.g. Oggiani (2021, 330), Oggiani (2022, 263) for bare nouns in Spanish.

(22) a. *El passaport donava permís per anar arreu d'Europa i*
the passport gave permission to go over of_Europe and
Amèrica.
America
'The passport gave permission to travel all over Europe and America'
(CTILC, 2013)

11 In some of the *give*-LVCs that are ditransitive, the indirect object (IO) would be introduced through an applicative head ApplP, as standardly assumed since Pylkkänen (2008).

b. **Permís era donat pel passaport per anar arreu d'Europa i*
permission was given for.the passport to go over of=Europe and
Amèrica.
America
'Permission was given by the passport to travel all over Europe and America.'

c. *El passaport dona permís per anar arreu d'Europa i*
The passport gives permission to go over of_Europe and
Amèrica, però només en/ #el donava als ciutadans majors
America but only it.PART/ it.ACC gave to.the citizens old
d'edat.
of_age
'The passport gave permission to travel all over Europe and America, but gave it only to citizens of legal age.'

Some crucial differences are found between bare singular nouns and bare plurals, as the latter do not present restrictions in terms of argument properties nor modification. This difference can also be traced back to Cyrino & Espinal's (2019, 192) proposal that bare plurals have a D layer in their syntax which is introduced by the plural marker. Therefore, bare plurals pattern with Type 1 LVCs in terms of determination and flexible modification, as expected.

In the remainder of the section, we focus on the patterns found regarding bare singulars in *give*-LVCs. First, they present some restrictions related to the type of modification and the verbal base of the noun. That is, when the noun selected by the LV *give* is deverbal and bare, modification is acceptable, though not especially frequent, with both relational adjectives (23) and qualitative adjectives (24).

(23) a. *Hay que darle respuesta afirmativa a este interrogante*.
have.IMP that give.it.DAT answer affirmative to this question
'One must give positive answer to this question'
(Spanish, CORPES XXI, 2001)

b. *una entitat catòlica que donava suport mèdic i financer a*
a entity Catholic that gave support medical and financial to
tèxtitles dones.
the women
'a Catholic entity that gave medical and financial support to the women'
(Catalan, CTILC, 2006)

c. *didn't really ask you here to give me spiritual advice*
(English, COCA, 2019)

(24) a. *no es posible dar respuesta precisa aún a ciertas preguntas*
not is possible give answer precise still to certain questions
'it is still not possible to give precise answer to certain questions'
(Spanish, CORPES XXI, 2002)

b. *A causa de la facilitat de fabricació i d'utilització, la prellosa dóna resposta eficient a la demanda de versatilitat funcional i arquitectònica del sostre.*
at cause of the facility of fabrication and use the pre-slab gives answer efficient to the demand of versatility functional and architectonical of.the ceiling
'Due to the ease in production and use, the pre-slab gives efficient answer to the demand of functional and architectural versatility in the ceiling.'
(Catalan, CTILC, 2001)

c. *I will do my best to give brief and unbiased advice.*
(English, COCA, 2012)

d. *Er gab später folgende Beschreibung*:
ge gave later following description
'He later gave (the) following description'
(German, DWDS, 2004)

However, non-deverbal bare nominals in *give*-LVCs exhibit certain cross-linguistic differences regarding modification. On the one hand, with non-deverbal nouns, Spanish and Catalan *give*-LVCs allow modification by a relational adjective, but reject qualitative modification, as shown in (25). In contrast, neither Germanic language includes modified bare singular non-deverbal nouns in our sample.

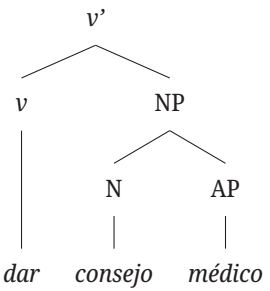
(25) a. *El profesor dio clase* {*particular*/ *magistral*/ **especial*}
the teacher gave class particular master special
'The teacher gave a {private / master / special} class' (Spanish)

b. *Va donar consell* {*mèdic*/ *formal*/ **interessant*}
goes.AUX.PST give advice medical formal interesting
'S/he gave a {medical/ formal/ interesting} advice' (Catalan)

Interestingly, the two kinds of adjectives (qualitative and relational) have been claimed to occupy different syntactic positions within the nominal domain, which are based on their restrictions; namely, a more internal position of the relational adjectives and a more external for the qualitative adjective (e.g. Fábregas, 2017, 77; see also Bonet and Solà, 1986, 313, for Catalan; and Bosque and Picallo, 1996, 379, for Spanish).

Partially following Oggiani's (2021; 2022) analysis of bare nouns in Spanish, where the different kinds of adjectives must be inserted at different levels within the noun phrase, depending on their closeness to the nominal, we suggest that bare nouns in *give*-LVCs can thus not project D, as in (26).

(26) *dar consejo medico*
give advice medical
'give (a) medical advice'



5.3 Type 3: An exceptional case of incorporation

There is an exceptional case of an LVC that does not fit into the previous types. It is the case of *dar vuelta* (lit. 'give turn') in Rioplatense Spanish[12] which can be built with a second direct object,[13] *la cabeza* 'the head', as in (27a), and can also be pronominalized with the accusative pronoun *la* (27b).

12 Spanish spoken in el Río de la Plata (Argentina) has the tendency to select bare nouns in more contexts than other varieties of Spanish (Rinaldi, 2018; Oggiani, 2021, 2022).

13 It is important to note that this is not a case of accusative doubling (typical from Rioplatense Spanish, as analyzed in Zdrojewski and Sánchez, 2014, 169, Di Tullio et al., 2019, 219) because it can also pronominalise a masculine antecedent with *lo*, as in (i), which refers to *el partido* 'the match'.

(i) *Con uno menos, Estudiantes se lo dio vuelta a River.*
with one less, Estudiantes SE it.ACC gave.3SG turn to River
'With one less, Estudiantes turned it over to River.'
(https://www.ambito.com/deportes/river/con-uno-menos-estudiantes-se-lo-dio-vuelta-n5169973)

(27) a. *Para que un día un macho cualquiera te diera vuelta la cabeza como una veleta.*
for that a day a male any you.DAT give turn the head like a weathercock
'So that one day any man would turn your head around like a weathercock.' (Spanish, CORPES XXI, 2001)

b. *te la diera vuelta como una veleta.*
you.DAT it.ACC give turn like a weathercock
'would turn it around like a weathercock'

However, the noun *vuelta* cannot be pronominalized (28b) nor extracted in interrogatives (28c). This is proof that it is not a canonical argument of the LV (as pointed out by Galbarini, 2017, 14).

(28) a. *Juan dio vuelta la hoja = Juan giró la hoja*
Juan gave turn the page Juan turned the page
'Juan turned the page'

b. **Juan la dio la hoja*
Juan it.ACC gave the page
'Juan turned it, the page'

c. **¿Qué dio Juan la hoja*?
what gave Juan the page
'What did Juan do the page?'

A third key aspect is adjacency, as there is no possibility of inserting any element between the LV *dar* and the noun *vuelta*, as in (29).

(29) *Dio* {**ya*/ **súbitamente*} *vuelta el partido*
gave already suddenly turn the match
'S/he gave {already/ suddenly} a turn to the match'

According to Acedo Matellán and Pineda (2019, 206), in similar Basque instances which allow a second DO, as in (30), "a lexicalization process has taken place, so that the argument structure of the LVC includes a slot for a DO with a Theme-role". Structurally, there is no trace preventing the merger of a DO and the LVCs can take a DP as a complement which is assigned absolutive case (as pointed out in Martinez, 2015, 376–377).

(30) a. *Ume-a-k* *eskol-ak* *huts* *egin ditu.*
kid-DET.ABS-ERG class-ABS.PL failure do AUX.PRS.ERG.3SG.ABS.3PL
'The kid missed the classes.'

b. *Gizon-a-k* *berrion-a* *hots* *egin* *du.*
man-DET.ABS-ERG good_news-DET.ABS noise do AUX.PRS.ERG.3SG.ABS.3SG
'The man spread the good news.'

These properties are aligned with a case-manipulating type of incorporation (Mithun, 1984, 857): the case left by the incorporated argument (the noun *vuelta*) is available for the other argument, i.e. the oblique argument, which is then promoted. For Mithun (1984, 856), the noun loses the syntactic status as an argument of the clause, when it is incorporated by the verb, which then leaves a case position vacated that can be occupied by the oblique argument.[14] Going back to the specific case of *dar vuelta*, in an early description of this construction, Masullo (1996, 194) suggests that the noun *vuelta* lacks reference due to low transitivity and, hence, allows for the second object to occupy its position. Also, Alonso Ramos (2004, 246–250) defends that there is no need for incorporation in such constructions because it is a ditransitive construction with two objects: one with a defective direct object (*vuelta*) and a full direct object (*hoja*). Although double object constructions are not typical of present-day Spanish, they were attested in Medieval periods of the Romance languages.[15] Their main particularity is that there is no univocal correspondence between the morphological case and the syntactic function.

Interestingly, Pineda (2014, 242) and Pineda (2015, 91–92) defends that some Catalan and Spanish ditransitive constructions should be analyzed as double object constructions. In her analysis, the object with accusative marking is a true IO. Following Pylkkänen (2008), she argues that the possibility that the IO presents accusative marking is explained by the fact that the low Applicative is not the regular Romance low Appl but rather the Anglo-Saxon applicative found in English double object constructions, which does not assign dative case and, hence, assigns the only available case (accusative) by default.

Besides the assignation of accusative case to the oblique object, the analysis of *dar vuelta* should account for the fact that the noun remains in situ (*vuelta*) and does not incorporate as an unergative verbs. At the same time, it should explain the fact

14 As pointed out by an anonymous reviewer, this could potentially be analyzed as a case of pseudo-incorporation (Dayal, 2011; Espinal and McNally, 2011). We leave this for further research.

15 Matute Martínez (2012, 960) studies these constructions in the history of Spanish and shows how the LVC followed by a direct object is not a productive structure after the 13th Century.

that there is no possibility to introduce an element between the LV and the noun, as previously seen in (29).

Within the extended projection framework (Grimshaw, 2005), the fact that the noun does not receive the accusative marking can be explained through the internal structure of the noun: *vuelta* does not project over n, and thus it cannot be case marked. The other object (oblique IO) has the accusative case available and can take it through the Applicative head proposed by Pineda (2015, 90).

Spanish is a language that a priori does not fall into the group of polysynthetic languages that undergo nominal incorporation (a la Mithun, 1984; Baker, 1988), which is a morphosyntactic process where a noun integrates in a verb or preposition which selects it, and it creates a complex predicate. In this case, the incorporated noun is highly defective and does not have number, case or definiteness marking, thus showing important differences with other more regular cases of incorporation. As Verdecchia (2021, 5) argues, the analysis of Spanish bare nominals cannot be done in terms of incorporation, since there is no strict adjacency between the verb and the noun (as would be the case in LVCs from Type 2). However, this is not the case for *dar vuelta* + DO and, hence, an incorporation analysis should be applied here to explain this unique LVC.

6 Conclusions

The present study reports the findings of a corpus-based study of adjectival modification in *give*-LVCs in two Germanic and two Romance languages. Our results have shown that the frequency of modification in LVCs is only higher than 50% in English, while results of modification in German, Catalan and Spanish LVCs are around a third of all instances. Such a contrastive approach to modification has proven that there are certain differences between language families, and even within the Germanic languages.

Taking the empirical results into account, which show certain restrictions and interactions, we have proposed a typology of LVCs which focuses on the possibilities of determination and modification of the nominal element. The data from Germanic and Romance languages presented in this study confirm that the presence (or absence) of the determiner is crucial for the degree of cohesion between the LV and the noun, which also affects modification.

First, there are LVCs with a determined noun that takes flexible modification (Type 1). The analysis for such constructions within a neo-constructionist approach to argument structure is that of a transitive creation verb, where the verb is followed by a DP that behaves like a canonical argument (Acedo Matellán, 2016; Myler,

2016). Second, there are LVCs with bare nominals (Type 2). The analysis proposed for this type is that they cannot project the DP layer, because they are not argumental. The inability of such noun phrases to behave like canonical arguments in LVCs explains why LVCs have traditionally been analyzed as complex predicates (Butt, 2010) or pseudo-incorporated nominals (Massam, 2001; Espinal and McNally, 2011). However, there is no need for such an analysis of LVCs, because there is no actual adjacency between the LV and the noun. Finally, there is an exceptional case of a LVC (*dar vuelta*) restricted to Rioplatense Spanish where there is an extra direct object and strict adjacency between the LV and the noun *vuelta*.

Therefore, the present proposal defends that all LVCs behave like regular verb-noun constructions (in line with Bruening, 2015), because the differences in the degree of cohesion between the LV and the noun are in fact due to the nature of the nominal rather than the LV itself.

Finally, our findings raise relevant issues in relation to the syntactic analysis of nouns in general and in LVCs in particular. Specifically, the general tendency of English LVCs is to have modified nominals, which was also found in previous studies (i.e. Bonial and Pollard, 2020). This has previously been used to argue that the weight of the nominal is the main reason for the coexistence of LVCs and the corresponding synthetic verb. However, our corpus findings have proven that English LVCs considerably diverge from the behavior of LVCs in the other three languages: German, Catalan, and Spanish, and thus cast doubt on this largely accepted hypothesis. Thus, the coexistence of LVCs and their synthetic counterparts, more specifically their semantic correspondence, should be the focus of future investigation. This further shows that the literature on LVCs would benefit from more contrastive-focused studies which pay attention to the morphosyntactic properties of LVCs across languages.

Bibliography

Acedo Matellán, Víctor. 2010. *Argument structure and the syntax-morphology interface. A case study in Latin and other languages*. Doctoral dissertation: Universitat de Barcelona. http://hdl.handle.net/10803/21788, last access: 13/01/2025.

Acedo Matellán, Víctor. 2016. *The Morphosyntax of Transitions*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198733287.001.0001.

Acedo Matellán, Víctor and Anna Pineda. 2019. Light verb constructions in Basque and Romance. In Ane Berro, Beatriz Fernández and Jon Ortiz de Urbina (eds.), *Aligning grammars: Basque and Romance*, 176–220. Leiden: Brill. https://doi.org/10.1163/9789004395398_007.

Alonso Ramos, Margarita. 2004. *Las construcciones con verbos de apoyo*. Madrid: Visor Libros.

Alvarez-Morera, Georgina. 2023. *The nominal in light verb constructions: a corpus-based study in present-day English, German, Catalan and Spanish*. Doctoral dissertation: Universitat Rovira i Virgili, Tarragona.

Baker, Mark. 1988. Theta theory and the syntax of applicatives in Chichewa. *Natural Language & Linguistic Theory* 6. 353–389.

Bonet, Sebastià and Joan Solà. 1986. *Sintaxi generativa catalana*. Barcelona: Enciclopèdia Catalana.

Bonial, Claire and Kimberly A. Pollard. 2020. Choosing an event description: What a PropBank study reveals about the contrast between light verb constructions and counterpart synthetic verbs. *Journal of Linguistics* 56(3). 577–600. https://doi.org/10.1017/S0022226720000109.

Bosque, Ignacio and Carme Picallo. 1996. Postnominal adjectives in Spanish DPs. *Journal of linguistics* 32(2). 349–385. https://doi.org/10.1017/S0022226700015929.

Bratánková, Leontýna. 2013. Le costruzioni italiane a verbo supporto. Un'analisi condotta sul corpus parallelo ceco-italiano. *Acta Universitatis Carolinae Philologica* 2. 55–70.

Brezina, Vaclav. 2018. Lancaster Stats Tools online. http://corpora.lancs.ac.uk/stats, last access: 13/01/2025.

Bruening, Benjamin. 2015. Light verbs are just regular verbs. *U. Penn Working Papers in Linguistics* 21(1). 1–10.

Bruker, Astrid. 2011. *Entwurf und Realisierung von Lexikon-Einträgen für Funktionsverbgefüge der deutschen Sprache*. Hamburg: Diplomica.

Butt, Miriam. 2010. The light verb jungle: Still hacking away. In Mengistu Amberber, Bret Baker and Mark Harvey (eds.), *Complex Predicates in Cross-Linguistic Perspective*, 48–78. Cambridge: Cambridge University Press.

Carnie, Andrew. 2002. *Syntax: A Generative Introduction*. Oxford: Blackwell.

Cattell, Ray. 1984. *Composite Predicates in English*. Sydney: Academic Press.

Cyrino, Sonia and M. Teresa Espinal. 2019. On the Syntax of Number in Romance. *Studia Linguistica* 74. 165–203. https://doi.org/10.1111/stul.12123.

Dayal, Veneeta. 2011. Hindi pseudo-incorporation. *Natural Language & Linguistic Theory* 29(1). 123–167. https://doi.org/10.1007/s11049-011-9118-4.

De Miguel, Elena. 2011. En qué consiste ser verbo de apoyo. In M. Victoria Escandell Vidal, Manuel Leonetti and Cristina Sánchez López (eds.), *60 problemas de gramática dedicados a Ignacio Bosque*, 139–147. Madrid: Akal.

Di Tullio, Ángela, Andrés Saab and Pablo Zdrojewski. 2019. Clitic Doubling in a Doubling World: The Case of Argentinean Spanish Reconsidered. In Ángel Gallego (ed.), *The Syntactic Variation of Spanish Dialects*, 215–244. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780190634797.003.0008.

Egbert, Jesse, Tove Larsson and Douglas Biber. 2020. *Doing linguistics with a corpus: Methodological considerations for the everyday user*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108888790.

Elenbaas, Marion. 2013. The synchronic and diachronic status of English light verbs. *Linguistic Variation* 13(1). 48–80. https://doi.org/10.1075/lv.13.1.02ele.

Espinal, M. Teresa and Louise McNally. 2011. Bare nominals and incorporating verbs in Catalan and Spanish. *Journal of Linguistics* 47. 87–128. https://doi.org/10.1017/S0022226710000228.

Fleischhauer, Jens. 2021. Light Verb Constructions and Their Families – A Corpus Study on German '*stehen unter*'-LVCs. In *"Proceedings of the 17th Workshop on Multiword Expressions (MWE 2021)*, 63–69. Online: Association for Computational Linguistics. 10.18653/v1/2021.mwe-1.8. https://aclanthology.org/2021.mwe-1.8, last access: 13/01/2025.

Fleischhauer, Jens. 2022. *stehen unter*-Funktionsverbgefüge und ihre Familien. *Zeitschrift für germanistische Linguistik* 50(2). 247–288. https://doi.org/10.1515/zgl-2022-2055.

Fábregas, Antonio. 2017. The syntax and semantics of nominal modifiers in Spanish: interpretations, types and ordering facts. *Borealis. An International Journal of Hispanic Linguistics* 6(2). 1–102. https://doi.org/10.7557/1.6.2.4191.

Galbarini, Paola. 2017. Diferentes usos del verbo dar: análisis desde un modelo (neo)construccionista. *Revista Pilquen* 20(4). 1–17.

Ginebra, Jordi. 2003. Fraseologia, concurrències lèxiques i llengua estàndard. In Miquel Àngel Pradilla (ed.), *Identitat lingüística i estandardització*, 7–55. Valls Tarragona: Cossetània Edicions.

Ginebra, Jordi and Pere Navarro. 2015. Concurrències lèxiques en català i en espanyol: uns quants contrastos. In Àlex Martín Escribà, Adolf Piquer Vidal and Fernando Sánchez Miret (eds.), *Actes del Setzè Col • loqui Internacional de Llengua i Literatura Catalanes. Universitat de Salamanca, 1-6 de juliol de 2012. Vol. II*, 217–228. Montserrat: Publicacions de l'Abadia de Montserrat.

Gries, Stefan. 2023. Statistical methods in corpus linguistics. In Foluke O. Unuabonah, Rotimi O. Oladipupo and Florence Daniel (eds.), *Readings in corpus linguistics: A teaching and research guide for scholars in Nigeria and beyond*, 78–114. Ibadan: Kraft Books Ltd.

Grimshaw, Jane. 1990. *Argument Structure*. Cambridge, MA: The MIT Press.

Grimshaw, Jane. 2005. *Words and Structure*. Stanford: CSLI Publications.

Grimshaw, Jane and Armin Mester. 1988. Light verbs and θ-marking. *Linguistic Inquiry* 19. 205–232.

Hale, Ken and Samuel Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: The MIT Press.

Hale, Ken and Samuel Keyser. 53–109. On argument structure and the lexical expression of syntactic relations. In Ken Hale and Samuel Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge, MA: The MIT Press.

Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Jespersen, Otto. 1942. *A Modern English Grammar on Historical Principles, Part VI, Morphology*. Copenhagen: Ejnar Munksgaard.

Leech, Geoffrey, Marianne Hundt, Christian Mair and Nicholas Smith. 2009. *Change in Contemporary English. A Grammatical Study*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511642210.

Levin, Magnus and Jenny Ström Herold. 2015. Give and Take: A contrastive study of light verb constructions in English, German and Swedish. In Signe Oksefjell Ebeling and Hilde Hasselgård (eds.), *Cross-Linguistic Perspectives on Verb Constructions*, 144–168. Newcastle upon Tyne: Cambridge Scholars Publishing.

Longobardi, Giuseppe. 2005. Toward a Unified Grammar of Reference. *Zeitschrift für Sprachwissenschaft* 24. 5–44. https://doi.org/10.1515/zfsw.2005.24.1.5.

Martinez, Arantzazu. 2015. *[Izen + egin]a aditz-lokuzioak: inkorporazio mailak*. Doctoral dissertation: Deustuko Uniberstitatea, Bilbao.

Massam, Diane. 2001. Pseudo noun incorporation in Niuean. *Natural Language & Linguistic Theory* 19. 153–197. https://doi.org/10.1023/A:1006465130442.

Masullo, Pascual José. 1996. Los sintagmas nominales sin determinante: una propuesta incorporacionista. In Ignacio Bosque (ed.), *El sustantivo sin determinación: la ausencia del determinante en la lengua espanola*, 169–200. Madrid: Visor.

Matute Martínez, Cristina. 2012. Aspectos de variación de las locuciones verbales con sustantivos escuetos en la historia del españo. In E. Montero Cartelle and C. Manzano Rovira (eds.), *Actas del*

VIII Congreso Internacional de Historia de la Lengua Española, I, Santiago de Compostela, 959–970. Santiago de Compostela: Meubook.
Mendívil, José-Luis. 1999. *Las palabras disgregadas: sintaxis de las expresiones idiomáticas y los predicados complejos*. Zaragoza: Prensas Universitarias de Zaragoza.
Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894. 10.1353/lan.1984.0038.
Myler, Neil. 2016. *Building and interpreting possession sentences*. Cambridge, MA: The MIT Press. https://doi.org/10.7551/mitpress/9780262034913.001.0001.
Oggiani, Carolina. 2021. "escribir artículo": nombres singulares escuetos en posición de objeto en español rioplatense. *Borealis. An International Journal of Hispanic Linguistics* 10(2). 313–333. https://doi.org/10.7557/1.10.2.5782.
Oggiani, Carolina. 2022. Los escuetos definidos débiles en español rioplatense / Bare weak definites in Rioplatense Spanish. *Revista de Estudos da Linguagem* 30(1). 239–269. https://doi.org/10.17851/2237-2083.30.1.239-268.
Piera, Carlos and Soledad Varela. 1999. Relaciones entre morfología y sintaxis. In Violeta Demonte and Ignacio Bosque (eds.), *Gramática descriptiva de la lengua española, vol. 3*, 4367–4422. Madrid: Espasa Calpe.
Pineda, Anna. 2014. *Les fronteres de la (in) transitivitat. Estudi dels aplicatius en llengües romàniques i basc*. Doctoral dissertation: Universitat Autònoma de Barcelona. http://hdl.handle.net/10803/283926, last access: 13/01/2025.
Pineda, Anna. 2015. Del Datiu a l'acusatiu. Un canvi sintàctic en procés en llengües romàniques i basc. *Llengua i literatura* 25. 73–98. Url=https://raco.cat/index.php/LlenguaLiteratura/article/view/298801, last access: 13/01/2025.
Pylkkänen, Liina. 2008. *Introducing arguments*. Cambridge, MA: The MIT Press.
Rácz, Anita, István Nagy and Veronika Vincze. 2014. 4FX: Light verb constructions in a multilingual parallel corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, 710–715. Reykjavik: European Language Association (ELRA).
Ramchand, Gillian. 2014. On structural meaning vs conceptual meaning in verb semantics. *Linguistic Analysis* 39(1-2). 207–243.
Rinaldi, Melisa Gisele. 2018. *Bare Singulars and So-Called Bare Singulars*. Doctoral dissertation: Queen Mary University, London. https://qmro.qmul.ac.uk/xmlui/handle/123456789/56405, last access: 13/01/2025.
Sanromán Vilas, Begoña. 2009. Diferencias semánticas entre construcciones con verbo de apoyo y sus correlatos verbales simples. *ELUA: Estudios Lingüísticos de la Universidad de Alicante* 23. 289–314.
Storrer, Angelika. 2007. Corpus-based investigations on German support verb constructions. In Christiane Fellbaum (ed.), *Collocations and Idioms: Linguistic, lexicographic, and computational aspects*, 164–187. London: Continuum Press.
Sánchez Rufat, Anna. 2016. El uso transitivo y ditransitivo de dar en un corpus escrito contrastivo. *Revista Fuentes* 18(2). 117–133. https://revistascientificas.us.es/index.php/fuentes/article/view/2917, last access: 13/01/2025.
Van Valin, Robert D. 2001. *An introduction to syntax*. Cambridge: Cambridge University Press.
Verdecchia, Matías. 2021. Nombres escuetos en contextos no verídicos. *Verba: Anuario Galego de Filoloxía* 48. 2–24. https://doi.org/10.15304/verba.48.6109.
Wierzbicka, Anna. 1982. Why can you have a drink when you can't *have an eat? *Language* 58(4). 753–799. https://doi.org/10.2307/413956.
Wiskandt, Niklas. this volume. Light verb constructions with experiencer objects in Germanic and Romance – A corpus-based contrastive perspective on 'make' and 'give' patterns. In Anna Riccio

and Jens Fleischhauer (eds.), *Light verbs: Synchronic and diachronic studies*, 97–122. Düsseldorf: Düsseldorf University Press.

Zdrojewski, Pablo and Liliana Sánchez. 2014. Variation in accusative clitic doubling across three Spanish dialects. *Lingua* 151. 162–176. 10.1016/j.lingua.2014.08.003.