THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH

Georgina Alvarez Morera

UNIVERSITAT ROVIRA i VIRGILI

The nominal in light verb constructions: a corpus-based study in present-day English, German, Catalan and Spanish

_____

Georgina Alvarez Morera

DOCTORAL THESIS

2023

The nominal in light verb constructions: a corpus-based study in present-day English, German, Catalan and Spanish

Georgina Alvarez Morera

Supervised by

Dr. Isabel Oltra-Massuet

Dr. Jordi Ginebra

DOCTORAL THESIS

Department of English and German Studies

Tarragona, 2023

UNIVERSITAT ROVIRA i VIRGILI

I state that the present study, entitled "The nominal in light verb constructions: a corpus-based study in present-day English, German, Catalan and Spanish", presented by Georgina Alvarez Morera for the award of the degree of Doctor, has been carried under my supervision at the Department of English and German Studies of this university.

Faig constar que aquest treball, titulat "El nom a les construccions amb verb de suport: un estudi de corpus en anglès, alemany, català i espanyol actuals", presentat per Georgina Alvarez Morera per a l'obtenció del títol de Doctor, ha estat realitzat sota la meva direcció al Departament d'Estudis Anglesos i Alemanys d'aquesta universitat.

Hago constar que el presente Trabajo, titulado "El sustantivo en las Construcciones con verbo de apoyo: un estudio de corpus en inglés, alemán, catalán y español actuales", presentado por Georgina Alvarez Morera para la obtención del título de Doctor, ha sido realizado bajo mi dirección en el Departamento de Estudios Ingleses y Alemanes de esta universidad.

MARIA ISABEL OLTRA MASSUET - DNI 38805191M
Digitally signed by MARIA ISABEL OLTRA MASSUET - DNI 38805191M
Date: 2023.08.17 19:35:53 +02'00'

Firmado por GINEBRA SERRABOU, JORDI (FIRMA) el día 17/08/2023 con un certificado emitido por AC DNIE 005

Isabel Oltra-Massuet

Jordi Ginebra Serrabou

Tarragona, 17 agost 2023

The Wish

Remember that time you made the wish?

I make a lot of wishes.

The time I lied to you
about the butterfly. I always wondered
what you wished for.

What do you think I wished for?

I don't know. That I'd come back,
that we'd somehow be together in the end.

I wished for what I always wish for.
I wished for another poem.

Meadowlands, Louise Glück

Acknowledgements

I started this journey on January 2020 and, hence, this dissertation has been mostly written in the solitude of two apartments. Fortunately, during this three and a half years I could connect with many people, mostly online. This thesis would not have been possible without them being at the other end of the internet.

First, I would like to thank with two supervisors, Isabel Oltra-Massuet and Jordi Ginebra. Their guidance, support and patience has led me to finish this work, even if at moments I was not confident enough. The discussion with my supervisors and the rest of the members of the ROLLING research group have helped me a lot during the process of building this project.

For the long coffees, I would like to thank especially my office colleagues: Celia Fullana, Vittoria Dentella and Camilla Masullo. *Grazie mille!*

I am also indebted to the two students of the Master's in advanced studies in Catalan, Yaiza de la Cruz and Edgar Batet, for their help in the retrieval and tagging of the Catalan database. To all the native speakers who were asked for acceptability of tests repeatedly, a massive thank you: Andrew, Shonaugh, Daniela, Felix and Catherine. *Vielen Dank!*

Last year, I had the opportunity to visit Dr. Antonio Fábregas at the University of Tromsø – above the Article Circle, who welcomed me in Tromsø and was always available whenever I had questions. I also want to thank all the members of CASTLFish I met during the seminars, workshops and readings groups. The three months spent there are among the most fruitful for this project. From a more personal perspective, I must thank the inhabitants in Casa Bekk, especially for the movie nights, never ending conversations by the fire and the chasing of the northern lights. *Tusen takk!*

I also want to thank Dr. Sara Gómez Seibane, my Master's thesis supervisor, for being the person that accompanied me during the first steps in corpus research. Her patience, optimism and constructive feedback convinced me to start working on a PhD Thesis. *¡Muchas gracias!*

I am grateful of having had the opportunity to attend several conferences and workshops around Europe, where I received valuable feedback which has enriched the final version of this dissertation.

On a more personal note, I must thank the friends who have been present, even from afar: Anna, Helena and Aleix from the UK; Adriana from Paris; Laura and Irene from Germany; and Laia and Albert from Wellington. To my friends in Konstanz for being the inspiration to start this journey, especially the two Marias, Vivek, Tim and Felix. *Moltes gràcies!*

Finally, thanks are due to my family. To my parents, Ernest and Griselda, for believing in me and giving me all the books I wanted as a child (and not only). *Gràcies per donar-me ales per volar*. To my sister, Grisi, for always being there, even at the library.

To Víctor, for being my first reader and having agreed to walk by my side during this journey.

I want to dedicate this dissertation to my grandmother, Carme Matges, who did not have the opportunity to study but had a great love for language. She would have made a great linguist.

Cover image by Adriana D. Martín-Zamorano.

Abstract

This dissertation presents an empirical and theoretical examination of light verb constructions (LVCs) in four languages: English, German, Catalan and Spanish. Its final objective is to provide a unified proposal for a crosslinguistic analysis of the nominal in LVCs. With this aim, the dissertation is organised in two main sections: a corpus-based study and a theoretical proposal.

The LVCs selected for this dissertation are constructions containing two elements: the light verb (LV) and the non-verbal element (NVE), a nominal from which the meaning of the entire construction is derived. The study investigates three LVs (*give, take* and *make*) across Germanic (English and German – *geben, nehmen* and *machen*) and Romance languages (Catalan – *donar, prendre* and *fer*, and Spanish – *dar, tomar* and *hacer*). Taking a random sample of 21,334 instances from online annotated corpora within the time frame of the 21st Century as a basis, I review the morphosyntactic properties of the nominal element, with a focus on the determination and modification patterns.

After the quantitative analysis, I observe that the majority of LVCs in all four languages appear unmodified, although there are significant differences within and across the language families: modified NVEs are most frequent in English, while German has the lowest proportion. In turn, the two Romance languages present around a third of NVEs with modification.

In an attempt to analyse the relationship between LVCs and their synthetic verbal counterpart (e.g. *give an answer – to answer*), I further study the Catalan LVCs in the dataset. The corpus-based study confirms that the weight of the nominal is an essential variable: LVCs with bare unmodified NVEs are more likely to be interchangeable with a synthetic verb.

In terms of the degree of frequency of LVCs, the most frequent lemmas do not exhibit signs of lexicalization in any of the languages under study. Also, in the comparison between LVCs and their synthetic verbal counterpart, the most frequent Catalan LVCs have shown to be highly interchangeable. These findings reject the view that LVCs coexist with the synthetic verbs because there is no semantic correlation or the LVC has developed a specific lexicalized meaning.

These findings constitute the foundation for an analysis of LVCs within a neoconstructionist approach to argument structure (Acedo-Matellán, 2010, 2016, based

on Hale & Keyser, 1993, 2002). In the second part of this dissertation, I propose a typology that focuses on the possibilities of determination and modification within the nominal domain. 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, 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), my claim is that it is the nature of the syntactic nominal phrase that establishes the combinatorial patterns of the LVC as a whole.

Resum

Aquesta tesi presenta un estudi empíric i teòric de les construccions amb verb de suport (CVS) en quatre llengües: anglès, alemany, català i espanyol. El seu objectiu final és proporcionar una proposta interlingüística per l'anàlisi de l'element nominal en les CVS. Amb aquesta finalitat, la tesi es divideix en dues seccions principals: un estudi de corpus seguit d'una proposta teòrica.

Les CVS seleccionades per a aquesta tesi són construccions que contenen dos elements: el verb lleuger o de suport (VS) i l'element no verbal (ENV), un substantiu del qual es deriva el significat de tota la construcció. L'estudi investiga tres verbs en llengües germàniques (anglès – *give*, *make* i *take*, i alemany – *geben*, *nehmen* i *machen*) i llengües romàniques (català – *donar*, *prendre* i *fer*, i espanyol – *dar*, *tomar* i *hacer*). A partir d'una mostra aleatòria de 21.334 ocurrències de corpus en línia dins del marc del segle XXI, revisem les propietats morfosintàctiques de l'element nominal centrant-nos en els patrons de determinació i modificació.

Després de l'anàlisi quantitativa, hem observat que la majoria de CVS apareixen sense modificació en les quatre llengües, tot i que hi ha diferències significatives entre les famílies de llengües: els ENV modificats són més freqüents en anglès, mentre que l'alemany té la proporció més baixa. Així mateix, les dues llengües romàniques presenten al voltant d'un terç de noms amb modificació.

Amb l'objectiu d'analitzar la relació entre les CVS i els seus verbs sintètics equivalents (e.g. *donar una resposta – respondre*), hem analitzat el conjunt de les CVS del català. Aquesta extensió de l'estudi de corpus confirma que el pes del nom influeix en la correspondència entre les CVS i el verb sintètic. De fet, les CVS amb noms escarits i sense modificació són més propenses a ser intercanviables per un verb sintètic.

Pel que fa a la freqüència de les CVS, les més freqüents no mostren signes de lexicalització en cap de les llengües estudiades. A més, en la comparació entre les CVS i els verbs sintètics equivalents, les CVS del català més freqüents han demostrat ser altament intercanviables. Aquests resultats descarten la proposta que les CVS coexisteixen amb els verbs sintètics perquè no hi ha una correlació semàntica exacta o la CVS ha desenvolupat un significat lexicalitzat específic.

Aquests resultats constitueixen la base per l'anàlisi de les CVS dins d'un enfocament neoconstruccionista de l'estructura argumental (Acedo-Matellán, 2010, 2016, basat en Hale & Keyser, 1993, 2002). A la segona part d'aquesta tesi, proposem una

tipologia que se centra en les possibilitats de determinació i modificació dins del domini nominal. Els resultats apunten cap a una jerarquia composicional del domini nominal basada en la noció de projecció estesa de Grimshaw (2005): com més independent del verb és el nom, més estructura es projectarà dins del domini nominal. A diferència d'estudis anteriors que apunten al paper del verb per explicar les restriccions a l'element nominal (Mendívil, 1999), la nostra proposta és que l'estructura sintàctica de l'element nominal és la que estableix els patrons de combinació de la CVS en el seu conjunt.

Resumen

La tesis presenta un estudio empírico y teórico de las construcciones con verbo de apoyo (CVA) en cuatro lenguas: inglés, alemán, catalán y español. El objetivo final es proporcionar una propuesta interlingüística para analizar el elemento nominal en las CVA. Con este propósito, la tesis se organiza en dos secciones principales: un estudio de corpus y una propuesta teórica.

Las CVA seleccionadas para esta disertación son construcciones que contienen dos elementos: el verbo ligero o de apoyo (VA) y el elemento no verbal (ENV), un sustantivo del cual se deriva el significado de toda la construcción. El estudio investiga tres VA en lenguas germánicas (inglés – *give*, *make, take*, y alemán – *geben*, *nehmen*, *machen*) y lenguas románicas (catalán – *donar*, *prendre*, *fer*, y español – *dar*, *tomar*, *hacer*). Tomando una muestra aleatoria de 21.334 ocurrencias de corpus en línea en el marco del siglo XXI, hemos revisado las propiedades morfosintácticas del elemento nominal con foco en los patrones de determinación y modificación.

Después del análisis cuantitativo, hemos observado que la mayoría de las CVA aparecen sin modificación en las cuatro lenguas, aunque existen diferencias significativas entre las familias lingüísticas: los ENV modificados son más frecuentes en inglés, mientras que el alemán tiene la proporción más baja. A su vez, las dos lenguas románicas presentan alrededor de un tercio de los sustantivos con modificación.

Con el objetivo de analizar la relación entre las CVA y sus correspondientes verbos sintéticos (e.g. *dar una respuesta – responder*), hemos analizado las CVA del catalán. Esta extensión del estudio de corpus confirma que el peso del sustantivo influye en la correspondencia entre las CVA y el verbo sintético. De hecho, las CVA con sustantivos escuetos y sin modificación son más propensas a ser intercambiables por un verbo sintético.

En cuanto a la frecuencia de las CVA, las más frecuentes no muestran signos de lexicalización en ninguna de las lenguas estudiadas. Además, en la comparación entre las CVA y los verbos sintéticos correspondientes, las CVA más frecuentes del catalán han demostrado ser altamente intercambiables. Estos hallazgos descartan la propuesta de que las CVA coexisten con los verbos sintéticos porque no hay correlación semántica o la CVA ha desarrollado un significado lexicalizado específico.

Estos hallazgos constituyen la base para un análisis de las CVA dentro de un enfoque neoconstruccionista de la estructura argumental (Acedo-Matellán, 2010, 2016,

basado en Hale & Keyser, 1993, 2002). En la segunda parte de esta tesis, proponemos una tipología que se centra en las posibilidades de determinación y modificación dentro del dominio nominal. Los resultados apuntan hacia una jerarquía composicional del dominio nominal basada en la noción de proyección extendida de Grimshaw (2005): cuanto más independiente del verbo es el sustantivo, mayor estructura se proyecta dentro del dominio nominal. A diferencia de estudios anteriores que se centran en el papel del VA para explicar las restricciones en el elemento nominal (Mendívil, 1999), nuestra propuesta es que la naturaleza sintáctica del sustantivo es la que establece los patrones de combinación de la CVA en su conjunto.

Table of contents

List of tables

List of figures

List of abbreviations

1	1	First person
2	2	Second person
3	3	Third person
A	ABS	Absolutive case
	ACC	Accusative case
	Adj	Adjective
	Appl	Applicative
	ASNs	Argument structure nominalizations
	AUX	Auxiliary
B	BPl	Bare plural
	BS	Bare singular
C	Cat.	Catalan
	CL	Clitic
	COMPL	Complement
	CP	Complementizer Phrase
D	D	Determiner layer
	DAT	Dative
	DET	Determiner
	DO	Direct Object
	DP	Determiner Phrase
E	Eng.	English
	ERG	Ergative case
F	F	Feminine
	Freq	Frequency
	FUT	Future
G	Ger.	German
I	IO	Indirect Object
	IP	Inflection Phrase
L	L1	First language (or native language)
	L2	Second language (or additional language)
	LEX	Lexical
	Lit.	Literally
	LL	Loglikelihood
	LV	Light Verb
	LVC	Light Verb Construction
M	M	Masculine
	MOD	Modified
N	N	Noun
	NC	Noun Classifier
	NOM	Nominative
	NP	Noun Phrase
	Num	Number
	NumP	Numeral Phrase
	NVE	Non-verbal element
P	P	Preposition
	PART	Participle
	PARTV	Partitive

	PERF	Perfect
	PI	Pseudo-incorporation
	PL	Plural
	PP	Prepositional Phrase
	PRED	Predicate
	PRS	Present
Q	Q	Quantifier
R	Refl	Reflexive
	Rel	Relative clause
	RNs	Result nominalizations
	RQ	Research question
S	SG	Singular
	Sp.	Spanish
U	Unmod	Unmodified
V	V	Verb
	v	little v
	V2	Verb second
	VoiceP	Voice Phrase
	VP	Verb phrase

Chapter 1. Introduction

This introductory chapter presents the scope of the study in §1.1 its main aims (§1.1.1), the research motivation (§1.1.2), the research questions (§1.1.3), as well as the methodology that has been employed (§1.1.4). It also introduces the main theoretical assumptions in §1.2, and it develops the layout of the present dissertation in §1.3.

1.1. Scope of the study

Light verb constructions (LVCs) are structures with a light verb (LV) and a non-verbal element (NVE), usually a nominal element, where the meaning is derived from the NVE, and the verb is considered *light* because its meaning is faded (Kearns, 1988). These constructions can be found across languages. In this study, a cross-linguistic focus will be applied to LVCs in four languages: two Germanic languages, English (1a) and German (1b), and two Romance languages, Catalan (1c) and Spanish (1d):

(1) a. take a walk

    b. einen Spazierganz machen

    c. fer un passeig

    d. dar un paseo

The term *light verb* (LV) was first coined by Jespersen (1954) to refer to verbo-nominal constructions in English like *have a sigh* or *take a shower*, where a deverbal noun is combined with a verb with a weak meaning, in the sense that a verb like *have* in the LVC *have a sigh* has no meaning of possession as the lexical verb *have* has in e.g. *have a house*. After this initial definition, different proposals within various frameworks in the field of theoretical linguistics have examined these structures. 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 NVE (i.e. Wierzbicka, 1982; Cattell, 1984; Grimshaw & Mester, 1988). However, more recent studies define LVCs as structures that are headed by a verb that has an abstract semantic content within the construction, but still contributes to the argument and event structure of the entire LVC (Ginebra, 2003; Ramchand, 2014).

The verb can contribute to the semantics of the construction in different ways: aspectuality changes, constraints on complements or thematic roles, and its presence is also necessary to license the predication of the NVE.

1

First, the presence of the LV contributes aspectual nuances to LVCs cross-linguistically (Traugott, 1999: 246; Butt, 2010: 66). For instance, a focus on the intention and result of the action is found in most *give*-LVC (2a) as opposed to their synthetic counterpart, where the reading is not necessarily deliberate or intentional (2b).

(2) a. She gave an answer (*involuntarily)

   b. She answered, involuntarily.   (Google Books)

The choice of LV can also affect the thematic roles, especially the number of roles that need to be assigned. For instance, in (3a) the LV *take* involves a single participant *Olive* but the LV *give* in (3b) includes a causative agent *Tom* in the event structure (Wittenberg et al, 2014: 64). Hence, the LV does contribute its own argument structure to the whole LVC, as proposed by Ramchand (2014).

(3) a. Olive took a bath

   b. Tom gave Olive a bath

For Spanish LVCs, Alonso-Ramos (2004) claims that the diathesis of LVs is dependent on the valency of the NVE with which they combine, as seen in (4). The contrast between *tener miedo* ('be afraid', lit. 'have fear') and *tener esperanza* ('have hope') is that the former has the stimulus encoded in a dative (4a), whereas in (4b) the stimulus is a genitive PP.

(4) a. Rosa le tiene miedo a la oscuridad.

      Rosa DAT has fear   at the darkness

      'Rosa is afraid of darkness'

   b. Rosa tiene esperanza de un cambio.

      Rosa has   hope   of a change

      'Rosa hopes for a change'   (Acedo-Matellán & Pineda, 2019: 192)

Last, the LV licenses the predication of a NVE. It generally combines with eventive nouns which already bear their own argument and event structure (Alexiadou, 2001), as in (5). LVs have been said to combine exclusively with eventive nouns in LVCs (Kearns, 1988; De Miguel, 2008, 2011), but LVs can also combine with referential nouns which *a priori* do not have predicative force (6).

(5) a. take a walk, give a hug, make a decision.

   b. einen Spaziergang machen, eine Umarmung geben, eine Entscheidung treffen.

   c. dar un paseo, dar un abrazo, tomar una decisión.

      d. fer un passeig, fer una abraçada, prendre una decisió.

(6) a. give a possibility, take surveillance, make a mistake.

      b. *Anteil nehmen* ('take an interest'), *Photo machen* ('take a picture'), *Platz nehmen* ('take a seat', lit. 'take a place').

      c. *dar una conferencia* ('give a conference'), *hacer punto* ('do stitch'), *tomar consistencia* ('take consistency').

      d. *donar l'enhorabona* ('congratulate', lit. 'give the congratulations'), *fer un discurs* ('make a speech'), *prendre precaució* ('take precautions').

The common characteristic of NVEs in LVCs seems to be that they must have argument structure. According to Alonso-Ramos (2004: 142), the NVE needs to be predicative, and it must have at least one argument to share with the LV. However, how these arguments are shared is still a matter of debate (see Chapter 2).

Another relevant property of LVs is that they are among the most frequent verbs in different languages (Butt, 2010; De Miguel, 2011), although the processing of LVCs by speakers of the language seems to be complex. Several psycholinguistics' studies (Wittenberg & Piñango, 2011; Wittenberg et al., 2014, among others) present evidence that the processing of LVCs involve increased reaction times in comparison with (a) their non-light counterparts (i.e. *give a kiss* takes more time to be processed than *give a flower*), and (b) their synthetic counterparts (i.e. *give a kiss* compared to *kiss*), which was considered as evidence for increased processing costs. Other studies based on the acquisition of Greek by children with specific language impairment (Kambanaros et al., 2014) confirm that they do not use LVCs, but only general all-purpose verbs when they are unable to produce a specific lexical verb. These findings from studies based on the processing and acquisition of LVCs seem to point to a special behaviour of these constructions.

Apart from the idiosyncrasies of LVCs which have been studied from different theoretical perspectives, LVCs usually have a synthetic verb which *a priori* conveys the same meaning (Sanromán Vilas, 2009; Butt, 2010): Eng. *walk* for *take a walk* or Sp. *pasear* for *dar un paseo* 'give a walk'. For this reason, LVCs have been considered as periphrastic paraphrases of simple verbs, but this equivalence is not always present.

In the correlation between an LVC and a synthetic verb, there are four possibilities (Piera & Varela, 1999): (i) the synthetic counterpart is equivalent in both its morphology and semantics (7a), (ii) there is morphological affinity but no semantic equivalence (7b),

(iii) there is only semantic equivalence (7c), and (iv) there is no synthetic counterpart within the language (7d).

> (7) a. take a shower – to shower, *Antwort geben* (lit. 'give answer') - *antworten*
> ('answer'), *dar besos* (lit. 'give kisses') – *besar* ('kiss'), *fer una còpia* (lit. 'make a copy') - *copiar* ('copy').
> b. give voice – to voice ('express'), *hacer fiesta* (lit. 'do holiday') – *festejar*
> ('celebrate'), *donar propina* (lit. 'give tip') - *propinar* ('give/administer')
> c. make a mistake – to err, *Platz nehmen* (lit. 'take (a) place') - *setzen* ('sit'), *hacer*
> *punto* (lit. 'do stitch') – *tricotar* ('knit'), *donar un mastegot* (lit. 'give a slap') -
> *pegar* ('hit').
> d. take advantage, *Abstand nehmen* ('reject')*, hacer deporte* ('do sport'), *donar*
> *una conferència* ('give a conference')

Even in cases where there is a correlation between the LVC and its synthetic counterpart (as in 7a), the semantic equivalence between the pairs is not always complete. For Spanish, Sanromán Vilas (2009: 311) proves that LVCs and their synthetic counterparts are not interchangeable in many contexts, for instance, when the subject is non-human (8a) or in performative utterances (8b).

> (8) a.*el viento daba golpes en ellas.
> 'The wind hit them.'
> b. ?Les doy la orden de que guarden silencio.   (Sanromán Vilas, 2009: 303)
> 'I command you to keep silent.'

Probably the most striking question raised by the apparent co-existence of an LVC and its synthetic counterpart relates to the economy principle in language. Namely, if language aims to expend the least amount of effort for a particular purpose (Zipf, 1949; Martinet, 1955), are these LVCs violating this principle? Or do they provide a valid alternative to their corresponding synthetic verbs?

In order to answer these and other questions, an analysis of LVCs in two Germanic languages (English and German) and two Romance languages (Spanish and Catalan) will be presented in this dissertation. LVCs in such a crosslinguistic approach have not been addressed so far, more specifically in this language combination.

The empirical basis of this study are LVCs with an eventive reading, as the examples presented so far in this introduction (1)-(8). LVCs with a causative reading (9)

4

are not considered in this analysis, since they have been studied as causative constructions (especially in Romance languages, e.g. Alsina, 1996, Alba-Salas, 2004).

(9) a. Sp. *dar miedo* ('give fear'), *dar vergüenza* ('give embarrasment')*, dar una sorpresa* ('give a surprise')*, hacer ilusión* ('do excitement')*, hacer gracia* ('do fun')   (Alonso-Ramos, 2004: 110)

b. Cat. *fer por* ('do fear'), *fer vergonya* ('do embarrassment'), *fer/donar una sorpresa* ('do/give a surprise'), *fer il·lusió* ('do excitement'), *fer gràcia* ('do fun')

In (9), the function of the LV in causative LVCs differs from other LVCs with an eventive reading. Alba-Salas (2004) for Italian, in line with Abeille (1988) for French, proposes that there are different variants of the light verb *fare* ('do/make'), and these variants are selected by different types of nominals. In the case of causative *fare* (10), the experiencer indirect object (*Ali*) is always directly dependent on the verb, and not on the NVE *paura* ('fear')*,* as they can only be clefted as separated constituents (10b).

(10)   a. Mark *fa paura* a Ali.   (Alba-Salas, 2004: 299)

Mark does fear to Ali

'Mark frightens Ali'

b. *È [*paura a Ali*] che Mark fa, non [schifo a Sara].

is fear to Ali that Mark does not disgust to Sara

'Mark frightens Ali, he doesn't disgust Sara'

In contrast, in *fare* LVCs (as in 14) selecting an eventive noun, there is a possibility of a double analysis, where the PP *a Eva* ('to Eva') can be either analysed inside the projection of *telefonata* ('call'), as in (11a), or directly dependent of the LV *fare*, as in (11b)[1].

(11)   a. Monica farà [una telefonata [a Eva]]

b. Monica farà [una telefonata] [a Eva]

'Monica will make a phone call to Eva'

Likewise, Gross (1989) already points out that causative verbs (or operators) followed by a noun can be paraphrased with a synthetic verb (12a), in the same way as LVCs (12b), but they differ in their structure. For this author, the noun *complexe* ('complex') in (12a) has only one argument (the subject, *Luc*, is not an argument of the

---

[1] This double analysis proposal is, however, controversial and it will be further discussed in Chapter 2.

noun because it is introduced by the causative verb), whereas in (12b) the noun *autorisation* ('authorisation') has three arguments: *Luc, Max, jouer.*

(12)   a. Luc a donné des complexes à Max. / Luc a complexé Max.

'Luc gave Max a complex'

b. Luc a donné à Max l'autorisation de jouer. / Luc a autorisé Max à jouer.

'Luc gave Max the authorization to play. / Luc authorized Max to play'

The structural differences between the LVs used in causative LVCs and other LVCs with an eventive reading in Romance languages have not been identified in Germanic languages such as English and German. Due to the contrastive focus of this investigation, the analysis will be restricted to the second group of LVCs which are found in all four languages under study. Thus, I will not incorporate LVCs as in (8) and (9) in the corpus database for Spanish and Catalan.

Finally, it is worth mentioning that the use of the term *light verb* in LVCs, as understood here, is not equivalent to the Chomskyan light verb (Chomsky, 1995), which is a functional verb with no phonological representation, as pointed out by Lin (2001) and Acedo-Matellán & Pineda (2019). In Chomsky's framework, the light verb (v) is an empty placeholder in the syntactic structure: it heads a phase, it has phi-features, it is semantically light, and it has no phonetic form nor selectional properties as regular verbs. Thus, it is light because it is an empty position (Lin, 2001: 22).

In the present study, however, the light verbs under analysis are phonologically realized, and they occupy a position within the syntactic structure.

1.1.1. Aims

This dissertation explores LVCs in four languages: English, German, Catalan and Spanish, using data from corpora as empirical basis. Thus, the first goal is to identify and contrast the properties of LVCs in all four languages with a special focus on the properties of the nominal component and differences among the light verbs.

Most LVCs have a corresponding synthetic verb that apparently has the same meaning. However, it has been shown that there are certain differences. The question is whether speakers convey the same meaning with the synthetic verb, e.g. *to answer,* and the LVC construction, e.g. *to give an answer*. Thus, the second goal of this dissertation is to identify the syntactic, semantic and pragmatic differences between the LVC and its

synthetic counterpart. This is done on the basis of Catalan, a language for which no such study has been undertaken yet.

My corpus data supports Acedo-Matellán & Pineda's (2019) analysis that there is a continuum in the degree of cohesion between the light verb and the NVE, from transitive-like structures to fossilized LVCs with an incorporated noun. Thus, the third goal is to develop a partial typology of LVCs in the four languages under study, English, German, Spanish and Catalan, based on the regularities and differences in their semantic-syntactic structure found in my corpus database.

Finally, the fourth aim is to put forward a syntactic analysis for three of the most frequent LVs: *make*, *take* and *give,* which can be applied to the four languages under study. The analysis will be based on Hale & Keyser's (1993, 2002, and much related work) approach to argument structure, as further developed in Acedo-Matellán (2010, 2016) and Mateu & Acedo-Matellán (2012).

1.1.2. Research interest

LVCs are challenging for theoretical linguistics because they have some idiosyncratic syntactic, semantic and pragmatic characteristics. There is a tradition of investigating LVCs since the 1950s when Jespersen (1954) first mentioned these constructions in his *English grammar.* Many researchers have tackled this topic from different theoretical perspectives and have focused on different aspects related to LVCs and have taken different languages as empirical basis. For instance, several previous studies have focused on particular languages, such as Japanese (Matsumoto, 1996), Chinese (Shen, 2004; Tsou & Yip, 2020; Xu et al., 2022), Persian (Folli, Harley & Karimi, 2005; Shomoossi & Shomoossi, 2012; Fleischhauer & Neisani, 2020; Berenjian & Rahimian, 2021), Indo-Aryan (Butt, 1995; Butt & Lahiri, 2013; Vaidya & Wittenberg, 2020), Portuguese (Gonçalves et al, 2010; Coelho & Sousa, 2017; Choupina & Brito, 2018), Italian (Alba-Salas, 2002, 2004; Folli & Harley, 2013), Basque (Oyarçabal, 2006), and Irish (Nolan, 2014), among others.

In the languages that are the empirical testing ground of this study, LVCs have also been previously studied within different theoretical approaches. Most studies have focused on the general properties of LVCs in the English language (Kearns 1988; Grimshaw & Mester, 1988; Brugman, 2001; Tu & Roth, 2011; Elenbaas, 2013; Nagy, Vincze & Farkas, 2013; Bonial, 2014; Wittenberg & Snedeker, 2014; Plante, 2014; Chen,

Bonial & Palmer, 2015; Ronan & Schneider, 2015; Bruening, 2016, among others). Some recent studies based on corpora (Levin & Ström Herold, 2015; Nagy, Rácz & Vincze, 2020; Martínez Caro & Arús Hita, 2020) have shown that English LVCs present some characteristics which are not shared with other close languages, such as German and Spanish, suggesting that LVCs in English are less diverse with lower number of individual LVCs and also a shorter variety of LVs (Nagy, Rácz & Vincze, 2020: 4). Also, Levin & Ström Herold (2015: 20) show that English LVCs and Swedish LVCs behave more similarly than in comparison with German LVCs, which tend to present more idiosyncrasies (Levin & Ström Herold 2015: 23).

There are studies focusing on German LVCs (*Funktionsverbgefüge*) from the first descriptions by Porzig (1957) and Helbig (1984), and more recent studies based on corpora (Storrer, 2007; Fleischhauer & Gamerschlag, 2019; Fleischhauer, 2021, among others). For Spanish, LVCs have been widely discussed from a grammatical perspective within different approaches: Mendívil (1999), Alba-Salas (2002), Alonso-Ramos (2004), Bustos Plaza (2005), Sánchez Rufat (2016) and Galbarini (2017) focus on syntax; while De Miguel (2006, 2008, 2011), Sanromán Vilas (2009, 2013, 2014, 2017) focus on the semantics of the constructions. Finally, fewer studies have concentrated on LVCs in Catalan specifically (Montserrat, 2014), and most previous studies have provided a contrastive focus of Catalan LVCs with other languages, such as English (Ginebra, 2006, 2008), German (Castell, 2011; Schmid, 2016) and Spanish (Ginebra & Navarro, 2015; Romera Martín, 2017; De la Cruz, 2020, 2021).

These studies scratch the surface of a complex topic, where a multilingual approach could contribute to a better understanding of how these frequent complex structures function. Since most studies on LVCs have focused on the English language and have provided a great insight on how these constructions function from different perspectives, a corpus-based study is proposed here in an attempt to contrast previous findings with examples from corpora in English and other languages: Spanish, Catalan and German.

To my knowledge, a combination of these four languages has not yet been investigated, although they have been previously compared in different combinations: English, Spanish and German (Nagy et al., 2020), English & Spanish (Arús Hita & Martínez Caro, 2020), English and Catalan (Ginebra, 2006, 2008), Spanish and Catalan (Ginebra & Navarro, 2015; Romera Martín, 2017; De la Cruz, 2020, 2021), German and

Catalan (Castell, 2011; Schmid, 2016). All of them focus, however, on reduced amount of data reduced to specific genres, such as literary fiction in Ginebra (2006, 2008), Castell (2011), Schmid (2016), De la Cruz (2020, 2021), Arús Hita & Martínez Caro (2020) or legal texts (Nagy et al. 2020). Therefore, my corpus-based study aims to cover more general and representative data from various genres, which can further contribute to the discussion on these constructions.

Previous corpus-based studies on LVCs provide a through description of these constructions and point towards some of their main properties. A better understanding of LVCs and their properties at a theoretical level could also contribute to the study of these constructions from an applied perspective.

In fact, the study of LVCs is interesting because of their location on a continuum between free and fixed constructions (Colominas, 2001: 18). One of the main characteristics of LVCs is that they do not allow a compositional translation. It is not uncommon that the same noun selects for a different verb in each language, as in (13a), and even within a language (13b, c).

(13)   a. *fer un passeig* (Catalan), *dar una vuelta* (Spanish), *take a walk* (English)

         b. take a decision (British English), make a decision (American English)

         c. *fer un petó* (Central Catalan), *donar una besada* (Balearic Catalan)

For human translation, such differences should not create problems, but these divergences pose an important challenge for the development of automatic translation. With the aim to train natural language processing technologies for automatic translation, previous studies have used multilingual corpora (Racz et al., 2014; Bonial & Pollard, 2020, among others). Since these technologies are trained with data from corpora, it is relevant to study these constructions on the basis of corpus data.

Further, the fact that LVCs cannot be literally translated can pose problems in the process of learning an L2. Some studies (Durrant & Schmitt, 2009; Laufer & Waldman, 2010; Li & Schmitt, 2010; Yamashita & Jiang, 2010, among others) have pointed out that collocations and other constructions, such as LVCs, pose a special challenge for L2 learners. Some of the reasons of this have to do with: incongruency with verb selection in L1 or lack of obvious semantic connection between its parts (LV and NVE). The solution that has been proposed in the literature for the problem that L2 speakers have with LVCs is to instruct them with focus on the input and grammatical reflection (Pérez

Serrano, 2017; Mihajlovska, 2017; Nygård, 2021). In this case, a better understanding of LVCs from a crosslinguistic perspective could provide evidence that could serve to develop L2 learning materials with special focus on the structures that need more attention in a specific language according to the L1 of the learners.

1.1.3. Research questions

The ultimate goal of this thesis is to present a crosslinguistic study based on large corpora to shed new light on the similarities and divergences of LVCs in different languages, mainly Germanic languages (English and German) and Romance languages (Catalan and Spanish). Based on frequency and complexity of these constructions, this dissertation seeks to answer the following research questions, organised around the aims of this research:

RQ1. Are there any divergences in the selection of the NVE by the LV across languages?

a. If so, what are the differences in terms of determination and modification?

b. Does the absence of determination influence the structures within the NVE?

c. Is the degree of frequency of the LVC a further determining factor?

RQ2. To which extent LVCs and their corresponding synthetic counterpart convey the same meaning?

a. If there are differences, what are the syntactic, semantic and discourse differences between the LVCs and the corresponding synthetic verb?

RQ3. How can LVCs be structurally accounted for within a neo-constructionist approach to argument structure?

These questions will be addressed from a descriptive perspective and supported by corpus data.

1.1.4. Methodology

In order to answer the research questions stated in §1.1.3, this dissertation develops a comparative, crosslinguistic corpus-study to LVCs that serves as a basis for a theoretical approach to the internal structure and associated properties of LVCs. That is why corpus data is used as a basis to create a typology of LVCs which can hold cross-linguistically, while the data are also checked against native speakers' grammatical judgments and against information found in grammar, dictionaries and previous literature

on these constructions. As is common in the linguistics literature, syntactic and semantic tests are further applied to the instances found in corpora in order to develop a detailed description and analysis of the properties of LVCs and their component parts, the light verb and the nominal.

From the basic repertoire of light verbs which can be found cross-linguistically (Butt, 2010), a total of three verbs have been selected for this study: *make, give* and *take,* with their corresponding verb in each language, as exemplified in (14).

(14)   a. Eng. make a claim, give a kiss, take a picture

b. Ger. Sorgen machen, eine Antwort geben, die Rücksicht nehmen

'worries make, an answer give, the consideration take'

c. Cat. fer un comentari, donar un consell, prendre una decisió

'make a comment, give an advice, take a decision'

d. Sp. hacer un cambio, dar un paseo, tomar una determinación

'make a change, give a walk, take a determination'

Apart from their frequent use in all four languages under study, these three verbs have in common the kind of eventuality they select: dynamic events with a prominent agentive role (Colominas, 2001: 202), although they also display some differences in their argument structure, which will be commented on in upcoming chapters[2].

The methodological approach to the corpus-based study is synchronic, that is, the focus is on data from the present-day use of the languages. The time frame is set to the 21st Century, from year 2000 to 2020. Data samples were retrieved from the following online available corpora[3]:

a.   Corpus of Contemporary American English (COCA): https://www.english-corpora.org/coca/

b.   Das Digitale Wörterbuch der Deutschen Sprache (DWDS): https://www.dwds.de/r

c.   Corpus textual informatitzat de la llengua catalana (CTILC): https://ctilc.iec.cat/

---

[2] The frequent verb *have* has been excluded because it does not share dynamicity with the three selected verbs, and it has also already been recently analysed in detail (i.e. Myler, 2016).

[3] Further details on the corpora can be found in Chapter 3.

d. Corpus del Español del Siglo XXI (CORPES XXI): https://www.rae.es/banco-de-datos/corpes-xxi

The corpus sample was obtained by searching for the collocates of the LV and the nominal. The following tools available in each corpus were used to extract the collocates: *Collocates* page in COCA, *DiaCollo* section in DWDS, *Col·locacions* in CTILC, and *Coaparición* in CORPES XXI.

After manually excluding instances of non-light structures, for every LVC a sample of up to 200 occurrences were analysed. The LVCs were organised according to the degree of frequency in the corpora under the label frequency: high, medium, and low.

Also, all examples were classified according to three grammatical factors of the NVE, namely,

(i)   determination (yes/no), including determiners such as articles, demonstratives,   possessives, numerals and other quantifiers.
(ii) modification (different classes), including adjectival modification, noun classifiers, prepositional phrases, genitive and relative clauses.
(iii) number of the noun (singular/plural).

The analysis of each of the selected verbs (*give, take* and *make*) is conducted for all four languages, in order to draw contrastive conclusions. For this purpose, LVCs extracted from corpora are analysed in terms of frequency and complementation patterns, as in (i) - (iii). The final database consists of 21,334 instances.

The results are presented alongside descriptive statistics: percentages and proportions have been carried out via Ms Excel's Analyze Data Toolbox. Further inferential statistics (chi-square tests) have also been produced to facilitate the interpretability of the results: when the relation between groups is significant ($p < 0.001$), it shows that the difference is a reflection of variation, and not just due to chance. In fact, the chi-squared test is one of the most commonly used significance tests in corpus linguistics for the analysis of variance, that is, the measure of data dispersion in statistics (McEnery & Wilson, 2001). Further details on the use of statistics in the analysis of empirical data will be given in the data analysis section (§3.2.4).

Additionally, in an attempt to compare the behaviour of each language regarding their modification patterns, a logistic regression modelling is implemented in Chapter 3.

The purpose of the implementation of this test is to explore the extent to which the response variable (that is, modification) can be predicted or explained in relation to the other variables examined (that is, the predictor variables) (Speelman, 2014). In this study, the predictors are both grammatical variables (LV, number and determination) and more general variables (the different degrees of frequency).

These techniques (chi-square tests and logistic regression modelling) are also implemented in the analysis of the data in Chapter 4, which focuses on the Catalan LVCs dataset in relation to its interchangeability by a synthetic verbal counterpart. As mentioned above, the relation between the LVC and its corresponding synthetic counterpart in Catalan is missing in the literature, and thus this chapter aims to fill this gap.

After the exhaustive comparative corpus-study in the first empirical part of this dissertation, the second part of this dissertation applies the hypothetico-deductive methodology commonly used in linguistic theory (Chomsky, 1965; Mendívil-Giró, 2019a). Specifically, based on the empirical data examined in the first part and a review of previous theoretical studies, a typology of LVCs is put forward which is verified through grammatical tests and native speakers' judgements (Chapter 6).

1.2. Basic theoretical assumptions

The aim of this thesis is to structurally account for the idiosyncrasies of LVCs within a neo-constructionist approach to argument structure. This theoretical framework is based on the assumption that argument structure can be explained within a syntactic system of structural relations between heads and their arguments (Acedo-Matellán, 2016; building on Hale & Keyser, 1993, 2002; Marantz, 1997; Mateu, 2002; Mateu & Acedo-Matellán, 2012). This syntactic system is based on two sets of building blocks: functional heads and roots, where the latter carry the conceptual content (as in Mateu, 2002; Acedo-Matellán, 2016).

The proposal mainly builds on the analysis put forward by Hale & Keyser (1993, 2002) for unergative verbs, as later developed in Mateu & Acedo-Matellán (2012) and Acedo-Matellán (2010, 2016). In this line, unergative verbs are analysed as underlying transitive structure with a light verb. Hence, the difference between unergative verbs like *dance* and its corresponding LVC (*do a dance*) is the nature of the complement within the structure: in (15a) the complement is the root DANCE, whereas in (15b) the complement

is a determiner phrase (DP) in a transitive construction where the DP *a dance* behaves like a canonical argument. In both cases, the LV is introduced in the little v.

(15) a. Unergative *danced*   b. Transitive creation *did a dance*

```
        VoiceP                              VoiceP
       /      \                            /      \
     Sue     Voice'                      Sue     Voice'
             /    \                              /    \
          Voice   vP                          Voice   vP
                 /  \                                /  \
                v   DANCE                           v   a dance
```

(Acedo-Matellán, 2016: 31)

Within this approach, Myler's proposal (2016: 24) for a definition of a LVC is that "[a] LVC is one that contains a v but no root". This explains how in (15b) *Sue did a dance,* the LV contributes event structure (being the realization of v) but no conceptual semantics (encapsulated in the root). Hence, the DP *a dance* supplies the event.

Since this dissertation focuses on the NVE in LVCs, I assume Abney's (1987) DP-Hypothesis for the analysis of the nominal domain. The main assumption is that a Determiner Phrase (DP) is the maximal category projected by the determiner elements of the lexical head, which is the noun, as represented in (16).

(16)

```
        DP
       /  \
     Spec  D'
          /  \
         D   NP
            /  \
          Spec  N'
                |
                N
```
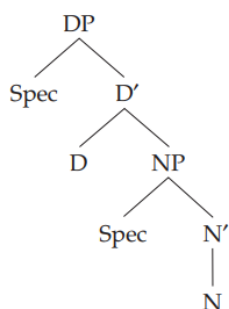
(Bernstein, 2001: 537)

Such an analysis has also been assumed for the analysis of noun phrases in the other three languages under study: German (Wiltschko, 1998), Spanish (Bosque & Gutiérrez-Rexach, 2009: §4.5; Leonetti, 2016; Fábregas, 2017), and Catalan (Cremades, 2014).

I assume, thus, that the syntax of LVCs can be studied with formal syntactic models that are already available such as the neoconstructionist approach to argument structure briefly presented in this section.

1.3. Layout

After the introduction, this dissertation is organised into two main parts: a corpus-based study (Part 1) and a theoretical proposal (Part 2).

First, the empirical section is divided into three chapters. Chapter 2 describes the properties of light verb constructions, as well as those of its component parts: the LV and the NVE. Chapter 3 presents the findings of the corpus-based study of the four languages under study (English, German, Catalan and Spanish) with focus on the modification possibilities in each light verb: *give, take* and *make* (and their corresponding verbs in all languages). Chapter 4 shows the results of the correspondence between LVCs and their synthetic counterpart taking the corpus dataset for Catalan LVCs as empirical basis.

Second, the theoretical section is divided into two chapters. Chapter 5 discusses the previous theoretical approaches to LVCs with special emphasis on the neoconstructionist proposals. Chapter 6 proposes a typology of LVCs in Germanic and Romance languages based on the cohesion between the LV and the NVE, which is itself based on the grammatical properties of LVCs studied in the previous chapters.

Finally, the conclusions are presented in Chapter 7 with a discussion of the main findings, the contribution of the thesis, as well as its limitations and areas for future research.

Chapter 2. Properties of light verb constructions

This chapter seeks to describe the general properties attributed to light verb constructions, as well as reviewing the contribution of previous corpus-based studies to the analysis of these constructions.

The first section describes the main properties of the LV itself in comparison to other verbal predicates in §2.1. First, the identification of the LV will be assessed with special focus on the form (§2.1.1). Then the contribution of the LV to the argument and event structure will be discussed (§2.1.2).

The second section focuses on the properties of the NVE within LVCs in §2.2.

The third section revolves around the interaction between the two elements in an LVC: LVs and NVEs in §2.3. First, the semantics will be discussed with a special focus on the predication function (§2.3.1). Then, some syntactic properties of LVCs will be described (§2.3.2). Finally, the degree of cohesion between the LVs and the NVEs will be discussed (§2.3.3).

The chapter concludes with a summary of the main properties of LVCs as described in the literature in §2.4.

2.1. Properties of the light verb

The term *light verb* has been widely used in English linguistics since Jespersen (1965, vol. VI: 167). Therefore, it is the term we adopt in this dissertation. However, these verbs have been referred to as *support verbs* (Gross, 1981; Piera & Valera, 1999; Alonso-Ramos, 2004; RAE-ASALE, 2009), *vicar verbs* (Mendívil Giró, 1999), and *empty verbs* (Mel'čuk, 1998), as well. While the terms *light* and *empty* focus on the semantic load of the verb and its weak contribution to the construction, both *support* and *vicar* highlight its role as syntactic support to the nominal element (Alonso-Ramos, 2004: 18). In all terminological traditions, there is general agreement that the verb does not determine the meaning of the construction.

In the remainder of this section, the identification of LVs will be discussed in relation to their form, and their similarities and differences with other verb types, specifically auxiliaries, serial verbs or full verbs (§2.1.1). After that, the contribution of

LVs to the whole LVC will be briefly introduced with a special focus on its argument structure (§2.2.2).

2.1.1. Identification of the LV: form

One of the main properties of light verbs (LVs) is that they have an identical form to a main lexical verb, which distinguishes them from auxiliary verbs, which have a defective paradigm in many languages, as in the Indo-Aryan language family studied in Butt & Lahiri (2002) and Butt (1995, 2010). That is, light verbs are like full verbs because they share their morphosyntax, i.e. they are form-identical to full verbs and show the same distribution, as well as their conjugational paradigm (Elenbaas, 2013: 50).

It is the semantic-syntactic environment that determines its use as a light or full verb, as illustrated in (1) and (2) for English and Spanish respectively. In (1a) and (2a), the semantics of the verb is bleached and the meaning of *give* and *take* is not literal (there is no actual action of *giving* or *taking*, but rather of *ordering* or *deciding*), whereas in (1b, 2b) the verb contributes its full semantic content.

(1) a. The captain gave an order to the crew. (LVC)

b. The captain gave a rose to the crew. (non-light)

(2) a. Rosa tomó la decisión antes. (LVC)

Rosa took.3SG the.SG decision before.

'Rosa made the decision before'

b. Rosa tomó los libros antes. (non-light)

Rosa took.3SG the.PL books before.

'Rosa took the books before'

It has also been noted that the basic repertoire of LVs is essentially the same across languages: *do/make*, *give*, *have*, *take*, etc. (Butt, 2010; Myler, 2016). In fact, Butt & Lahiri (2003) claim that there is a handful of verbs crosslinguistically (between 5 and 20) which can be used as either a main verb or an LV depending on the context.

Diachronically, there is always "a stable corresponding full or heavy version (of the LV) in all the languages in which they are found" (Ramchand, 2013: 217), unlike auxiliaries or other forms which have undergone grammaticalization over time. In other investigations by Butt (2003, 2010) and Butt & Lahiri (2013), LVs in Indo-Aryan languages are argued to be members of the verbal category and have not gone down the

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

grammaticalization cline. In their view, the full verb (i.e. *give, take* and *make*) and their corresponding light version are synchronic variants.

Building on Bowern's (2008) crosslinguistic analysis of the diachrony of complex predicates in Pacific, South Asian and Australian families, Butt & Lahiri's data show that there is no support for the grammaticalization (or reanalysis) of an auxiliary from a light verb. That is, the LV is not an intermediate step between the full verb and the auxiliary. Evidence for this claim is found in the diachrony of Indo-Aryan languages (Butt & Lahiri, 2013), where LVCs can be identified over thousands of years and are syntactically relatively stable.

This view introduced by Bowern and Butt & Lahiri with data from non-European languages is opposed to a grammaticalization analysis of LVs defended by Brinton & Traugott (2005), Brinton (2008) and Berlage (2012). They argue that English LVs have grammaticalized into functional words, more specifically into a grammatical marker of telic aspect. For instance, already in the later period of Middle English, LVCs were associated with countable activities and a difference in meaning arose between *bathe* (wash on any occasion) and *have a bath* (prepare a bath) (Brinton & Traugott, 2005: 130). This claim is, however, contradicted by other diachronic corpus-based studies on Middle English, as in Claridge (2000: 245-246), where the telicity marking of emerging LVs is found to be rare. Also, synchronic studies of English LVs (Bonial & Pollard, 2020) show that the aspectual marking of LVCs is not determined only by the LV, but rather the NVE plays an important role (as will be further discussed in 2.2).

However, although the morphology of these verbs presents some changes over time, the co-occurrence of an inflected LV with the NVE remains constant historically, as well as the choice of the verb[4]. Butt & Lahiri (2013) suggest that the stability of LVs is due to their passepartout function, that is, they have a variety of uses. More recent studies provide further evidence against the grammaticalization analysis, showing that LVs are synchronic variants of full verbs both in English (Elenbass, 2013: 73), and Romance languages (García Pérez, 2007; Montserrat, 2014), as well as Basque (Acedo-Matellán & Pineda, 2019).

---

[4] The selection of some verbs has changed over time due to different contextual and intra-linguistic reasons. For instance, Montserrat (2014) studies the preference of Catalan LV *fer* ('do/make') over *donar/dar* ('give') from the 13th Century and García Pérez (2007) analyses the opposite tendency in the history of the Spanish language for the LV *hacer* ('do/make').

Having identified the light uses of the verb, it is worth having a look at the specific properties of the LV itself first. Initially, when the term *light verb* was coined by Jespersen (1954), it was considered to be completely devoid of predicational force and, thus, the argument structure of the LVC was inherited from the nominal element (Grimshaw & Mester, 1988; Hopper & Traugott, 2003). However, some of the properties attributed to LVs have been later disproven.

One of the main controversies regarding LVs is their status, for which three main approaches have been identified: whether (i) they are a separate syntactic class (Butt, 2010), (ii) they conform a subclass of semi-lexical verbs (Mohanan, 2017) or (iii) they are regular verbs of the control-type (Bruening, 2016). Let us consider each one in turn.

For Butt (2010), LVs are a cohesive class whose properties distinguish them from auxiliary verbs and from serial verbs, as well as from main verbs as seen in (1-2). She provides crosslinguistic evidence to support her claim.

First, auxiliaries are a functional category; they have undergone a process of grammaticalization and contribute tense and aspect to the predication (Heine, 1993; Hopper & Traugott, 1993). In contrast, LVs contribute to the semantics of the construction in different ways. For instance, they may add aspectuality, as proven by the difference between the verb *to answer* and the LVC *to give an answer* in (3), where the LVC involves agentivity and volitionality of the subject (3b).

(3) a. John *answered* involuntarily

b. John *gave an answer* (*involuntarily)

In addition, in other languages, the LV can require some constraints on the arguments, such as the thematic role marking in (4), where the Urdu LV *li* (equivalent to English *take*) marks ergative case on the subject (4a) but non-agentive LV *pAr* (equivalent to English *fall*) does not (4b).

(4) a. nadya=ne ro li-ya

Nadya.F.SG=ERG cry take-PERF.M.SG

'Nadya cried (has finished and did it on purpose).'   (Urdu in Butt, 2010)

b. nadya ro pAr.-i

Nadya.F.SG.NOM cry fall-PERF.F.SG

'Nadya fell to crying (involuntarily).'   (Urdu in Butt, 2010)

19

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

Taking into account that LVs contribute to the argument structure of the whole construction, Butt (2010) defends that these verbs are not completely semantically empty, although the degree of lightness is difficult to measure.

Second, LVs are also different from serial verbs, which are found in several languages, especially in West African, Oceanic and Asian languages (Seiss, 2009). In fact, there is consensus that auxiliary verbs develop from main verbs and acquire functional properties, whereas serial verbs lie in an intermediate stage on the grammaticalization cline (p. 506). For instance, Ngan'gityemerri has developed a serialized posture verb construction (as analysed by Reid, 2002), where a posture verb cliticizes onto an inflecting verb and a coverb[5] complex and adds aspectual information, as in (5).

(5) Inflecting Verb + Coverb + Encliticized Inflecting Verb:

Nganni-batybity-tye-*nginni.*

1Plex-Spoke.pi-sew.Past-1Plex.sit.pi

'We were sewing.'   (Reid, 2002: 256)

However, Seiss (2009) argues that such constructions should not be considered serial verb constructions, but rather the serial verb should be analysed as an auxiliary. As stated before, they present a different behaviour from LVs which have a clearer contribution to LVCs (as shown in 3-4).

Also, in languages like Jaminjung, a Northern Australian language, the LV combines with a coverb, as in (6), where the LV *hit,* in combination with the coverb *appear,* adds the aspectuality "suddenly", which changes the Aktionsart of the joint predication.

(6) a. bul ga-ruma-ny

   appear 3SG-come-PAST

   'appear'   (Jaminjung in Schultze-Berndt, 2002)

   b. bul   gani-ma

   appear 3Sg-hit-Past

   'appear (suddenly)'   (Jaminjung in Schultze-Berndt, 2002)

---

[5] Coverbs appear alongside inflected verbs in certain constructions in Australian, East African, Iranian, and Oceanic languages. Unlike inflected verb, they are not finite. Although coverbs are predicational and are not derived from another part of speech, coverbs have been analysed as a distinct part of speech class (Amberber, Baker & Harvey, 2007, 2010).

In sum, the contribution of the LV to the aspect of the predication is the main property that sets them apart from auxiliaries and serial verbs. For Butt (2010), this crosslinguistic evidence is the proof that LVs constitute a separate syntactic class. This is, however, still under debate.

Building on Butt's work, Mohanan (2017) considers LVs as a subclass of semi-lexical verbs, which are neither fully lexical nor fully grammatical (Corver & Van Riemsdijk, 2001), so she groups them into *grammatical verbs* because they are a closed class:

> Likewise, grammatical verbs refer to a special closed classes of verbs that do not permit addition of new members, except perhaps through the process of grammaticalization. Grammatical verbs typically include *auxiliaries*, indicating relative time, perfectivity, or progression; *modals*, indicating meanings such as those of obligation or permission; special verbs of possession (eg. *have*) and existence (eg. *be*); and *co-verbs*, so called especially in the literature on East Asian languages. (Mohanan, 2017: 2)

The main difference between LVs and the rest of the verbs included in this group of grammatical verbs is that the LV contributes to the argument structure of the construction jointly with the NVE (Mohanan, 2017: 5).

Building on Butt's generalization, Ramchand (2013: 7) discusses the structural and conceptual meaning of these verbs: "If light verbs were functional elements and their heavy counterparts were roots, there would be no ready explanation of the fact that certain types of meaning (event structural) carry over from one use to the other." After comparing LVs with their heavy versions cross-linguistically, the main conclusion is that both light and heavy uses of the same verb are systematically related, and what distinguishes LVs from full verbs and auxiliaries is their argument and event structure.

After all, if the LV is a main verb that is used in its light form in certain contexts, how is this represented in its lexical entry? Myler's (2016) proposal for the verb *have* is that its lexical entry is meaningless in order to account for all its uses, auxiliary, possession and also as a LV. His proposal aims for a unified analysis of all its uses, as "ultimately, more than one lexical entry for *have* or special stipulated conditions on interpretation turn out to be needed in order to make analyses work when they include the assumption that *have* has its own lexical semantics" (p. 175).

Although this does not need to be the case for all LVs, a proposal where the lexical semantics of the verb is empty, and it is filled up with the semantics of the constructions where it is inserted, could explain the many different contexts where these verbs can be found.

In contrast, Bruening (2016) focuses on LVCs in English and he argues that LVs are just regular verbs with little semantic specification: "[w]e only need two things to account for light verbs, and these are independently needed: the fact that NPs can be eventive; and the acknowledgement that the logical arguments of NPs can be controlled" (p. 52). The first argument is the fact that the argument structure of LVCs is determined by the verb. In the contrast between *give* and *take* selecting the same noun *punch*, the first verb allows an IO (7a) and the second only allows a PP (7b). In fact, *take* as a LV is not a verb of physical transfer, that's why the IO is not permitted (Norvig & Lakoff, 1987).

(7) a. She gave him a punch. / *She gave a punch at him.

b. She took a punch at him. / *She took him a punch.   (Bruening, 2016: 56)

If the argument structure was determined by the NVE, the instances in (7) should have the same argument selection. The LV is, thus, the contributor of the argument structure of the whole construction.

In sum, Table 1 includes a comparison between LVs and other verb types, as described in previous studies. The first two formal characteristics set them apart from auxiliaries and serial verbs. These two verb types present a defective paradigm and signs of grammaticalization over time, while LVs do not. It is the semantic characteristic what distinguishes LVs and full verbs, as they are identical in form.

Verb type	Defective paradigm	Grammaticalization	Semantics
Auxiliaries	Yes	Yes	Functional
Serial verbs	Yes	Yes	Functional
LVs	No	No	Light
Full verbs	No	No	Heavy

Table 1. Summary: Comparison of LVs properties versus other verb types

2.1.2. Contribution of the LV to the argument structure

The contribution of the LV to the argument structure of the construction has been an ongoing debate since Jespersen first coined the term *light verb* (1954) and it is strictly related to the very notion of *lightness*. Acedo-Matellán & Pineda's (2019: 178-192)

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

highlight that the general tendency in the 70s and 80s was to consider that the LV was completely light and only contributed the verbhood, while the LVC inherited the argument structure from the NVE (as in Grimshaw & Mester, 1988, for Japanases *suru* 'do', or Cattell, 1984, for English LVs). A different analysis is proposed within a lexicalist framework where there is a distinction between the lexical conceptual structure from which the LV is delinked and the syntactic argument structure which is provided by the LV with a particular number of argument slots, as well as event structure (Kearns, 1988).

Later, Butt's (1995) lexical-functional grammar analysis proposed the *argument fusion* operation which implies that the LV and the NVE become a single predicate. In her approach, LVCs are complex predicates syntactically formed through argument fusion, as opposed to lexically formed complex predicates where the meanings of both elements are merged. This is illustrative in the permissive structure in Urdu with the LV *de* 'give': in (8a), the LV contributes an argument (the permitter) in the joint predication; while in (8b), there is a biclausal construction with an embedded clause.

(8) a. nadya=ne yAssin=ko pAoda kAt . -ne di-ya

Nadya.F.Sg=Erg Yassin=Inst plant.M.Nom cut-Inf give-Perf.M.Sg

'Nadya let Yassin cut the plant.'　　　　　　　　(Urdu in Butt, 2010: 4)

b. nadya=ne yAssin=ko pAoda kAt . -ne di-ya

Nadya.F.Sg=Erg Yassin=Inst plant.M.Nom cut-Inf give-Perf.M.Sg

'Nadya let Yassin cut the plant.'　　　　　　　　(Urdu in Butt, 2010: 6)

In LVCs, the complex predicational unit hosts arguments contributed by either of the two predicates, but also arguments that are the result of fusing the two components since the subject is shared by both the LV and the NVE.

Building on Butt's proposal, Ramchand (2013) defends that LVs have the argument and event structure of the corresponding full verbs, and the difference between both uses, as LVCs and as non-light verbs, is reduced to the nature of the NVE. She verbalizes the following light verb constraint: a LV is "[a] verb can be used as a light verb when all of its category features agree with some other verbal element in its complement domain. If two lexical items combine to lexicalize an event structure skeleton, then their conceptual contents must be able to unify without contradiction" (p. 15). For Ramchand, the semantics of the constructions prove that LVs have more weight than a pure functional

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

head. Although it is true that LVs have an impoverished lexical content, their content is not null.

Generally speaking, the part of meaning which is shared between both uses of the verbs is the association with particular subevental heads (e.g. inceptive vs. completive transition). In contrast, what is not found in the light uses of these verbs is the physical motion, transfer and possession (Ramchand, 2013: 16). With this in mind, she opens the possibility that motion, location and transfer in space are cognitive defaults, which constitute another dimension of meaning, as in Table 2.

DOMAINS OF CONFLATION II		
Syn-Sem	Cognitive Defaults	Lexicon
Cause Non change vs. change Non scalar vs. scalar change Multivariate vs. bivariate transition Source of scale Result of change	Caused positional transfer Locations: -Manner of change -Change of location	EVERYTHING ELSE

Table 2. Lexical domains in Ramchand (2013: 19)

The mental processing and representation of LVCs have also shown that their semantic structure clearly differs from other constructions (Wittenberg et al., 2014). This is supported by a series of recent psycholinguistics studies, which have found that LVs take longer to process than non-light verb structures with the same noun; they also take longer than the same verb in a non-light use in both German (Wittenberg & Piñango, 2011) and English (Wittenberg et al., 2014). Their findings point to an idiosyncrasy of LVCs, since in 75% of the trials, participants assigned an agent-patient mapping to *give*-LVCs despite the ditransitive syntax, which has three arguments: source-theme-goal (Wittenberg & Snedeker, 2014: 6). These results were later validated in Wittenberg et al. (2017)'s study with mouse-click and eye tracking data, where they show that speakers do not categorise *giving a kiss* LVCs as either two or three role events: they are placed in an intermediate position. These results suggest an automatic influence of syntactic argument structure on semantic interpretation and event construal, even in highly frequent constructions.

After all, the contribution of the LV to the argument structure of the whole construction is key. LVs also provide part of the skeletal meaning: aspect, argument and event structure. Let us consider how each of them has been treated in the literature.

With regard to aspect, since Wierzbicka (1982), it has been proposed that LVCs have the aspectual function of being the telic counterpart (*have a thought*) of atelic verbs (*to think*). But not all LVCs have a uniform behaviour as regards telicity (Alonso-Ramos, 2004). Bonial & Pollard (2020) have investigated the aspectual differences between the synthetic verbs and the corresponding LVCs in English, and they have found that determiners play a role in the interpretation of telicity. In fact, there are multiple possibilities regarding determination of NVEs: they may always occur with a determiner (9a), only optionally (9b), or never (9c).

(9) a. He made *(a) decision.

b. He made {*a/ (some)} progress.   (Bonial & Pollard, 2020: 16)

c. He took (*a/*the) part in a tournament.

It seems that the LV does not affect the telicity of the NVE: if the deverbal nominal (e.g. *a rest*) is bounded, the presence of the LV (*take*) does not change its telicity (Elenbaas, 2013: 64). Hence, "this suggests that the aspectual function must be attributed mainly to the indefinite nominal predicate", which goes against previous accounts which see LVs in English as aspectual markers (Brinton & Traugott, 2005). Proof that LVs do not influence the telicity of the NVE are examples with the same LV (*take*) appearing in both contexts: atelic (10a) and telic (10b).

(10)   a. He took a rest {for/*in} an hour.

b. He took a breath {for/in} five seconds.   (Elenbaas, 2013: 64)

Elenbaas (2013) concludes that the telicity of the event depends on the verb from which the NVE is derived: *to rest* (atelic) and *breathe* (a/telic). For Portuguese, Gonçalves et al. (2010) and Choupina & Brito (2018) also show that LVCs have the same aspectual value than the NVE, in (11).

(11)   a. Dar uma caminhada (=process)

'Give a walk'

b. Dar uma pintura (=finalized process)

'Give a paint'

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

b. Dar um espirro (=point)

'Give a sneeze'

This is also suggested by the results of a corpus-based study on English *make* and Spanish *hacer* (Perea Irigoyen & Róo Sánchez, 2013: 97), which also detect that the Aktionsart of the LVCs are determined by the NVE: if the NVE is limited by a determiner, the LVC is telic; whereas if the NVE has no limit (that is, a mass noun or a countable plural), the whole construction is atelic. However, it is not always the case that the deverbal noun always inherits the aspect of the verb from which it derives (contra Elenbaas, 2013). In (12a) the synthetic verb *llamar* ('call') has no ending point, but its deverbal noun (*un llamado*) in (1b) implies the end of the action, an achievement.

(12)   a. llamar a usuarios {por/*en} una hora   (atelic)

'Call users {for/*in} an hour'

b. hacer un llamado {*por/en} una hora   (telic)

'Give a call {for/*in} an hour'

In sum, it seems that there is proof from different languages that aspect and, more specifically, telicity of LVCs is directly influenced by the NVE and its nominal projection.

Another aspectual nuance attributed to the LV is volitionality. Butt (2010) argues that the LVC provides a volitional component. For instance, *give an answer* is necessarily interpreted as a deliberate action whereas the synthetic verb *to answer* does not have this interpretation, as in (13a,b). However, as seen in (13c), the LVC could also be accompanied with a modifier involving no volitionality: to *give an unintentional answer* in certain contexts, such as a Philosophy handbook.

(13)   a. He answered deliberately / He {unintentionally / accidentally} answered

b. Mary gave *a deliberate answer*.

c. as souls is beyond our arbitrary control, that it has to give *an unintentional answer* to the question and to the calling of our particular lives.   (Google Books)

In Sanromán Vilas' (2011) study on the pairs of Spanish LVCs and their equivalent single verbal counterpart, volitionality is also found to be part of the contribution of certain LVs. For instance, several Spanish LVs can co-occur with nouns expressing an action, but there are some aspectual differences which constrain their selection: *hacer*

'make' and *dar* 'give' with *cabezada* ('snooze') are neutral with respect to volitionality, whereas *echar* 'throw' adds volitionality (14).

(14)   Sin darse cuenta, dio / hizo /*echó una cabezada de la cual se levantó un
         poco sobresaltado...   (Sanromán Vilas, 2011: 259)
         'He took a snooze without noticing it and he woke up startled'

In sum, the fact that *hacer* 'make' and *dar* 'give' are more neutral can account for why they can be interchanged both within the same language (Sp. *hacer/dar masajes* 'make/give massages', Cat. *fer/donar una ullada* 'make/give a look') and between languages (Sp. *dar un paseo* 'give a walk', *fer un passeig* 'make a walk').

In a similar line, Nygård (2021) investigates the differences between the Spanish LVs *echar* 'throw' and *soltar* 'drop' and confirms Sanromán Vilas's (2011, 2017) thesis that the LV does contribute semantic content to the LVC, since the LV shares semantic features with their full verb base. In (15), the differences between *soltar/echar un discurso* 'drop/throw a speech' are based on the differences between these verbs: only with *soltar* (15a) the interpretation can be that of a spontaneous and/or improvised speech (*discurso*).

(15)   a. Pedro soltó ayer un discurso impresionante al público.
                                                              (Nygård, 2021: 13)
         Pedro dropped.3SG yesterday a speech impressive to the audience
         'Pedro gave an impressive speech yesterday to the audience'
         b. Allá Samper se echó un discurso de tarima   (Nygård, 2021: 108)
         There Samper SE threw.3SG a speech of stage
         'There Samper gave a stage speech'

Another aspect which has been attributed to the LV is certain constraints on the argument selection. There has been a special emphasis on the selection of the subject, as LVs are used in several languages to make a distinction between agentive and non-agentive actions: the difference between *do* and *come* in Urdu in combination with *memory,* as in (16); *dare* 'give' and *prendere* 'take' in Italian in combination with *sgridata* 'scolding', as in (17); or *give* and *get* in English with *the creeps* , as in (18). Although the subject is the same in (a) and (b), it is the selection of the LV which determines the agentivity of the external argument.

(16)   a. nadya=ne kAhani yad k-I   (Butt, 2010: 4)

Nadya.F.SG=ERG story.F.SG memory.F do-PERF.F.SG

'Nadya remembered the story.'

b. nadya=ko kAhani yad a-yi

Nadya.F.SG=DAT story.F.SG memory.F come-PERF.F.SG

'Nadya remembered the story (the memory of the story came to Nadya).'

(17)   a. Ha preso una sgridata   (Folli & Harley, 2013: 104)

Has.AUX taken.PART a reprimand.

'He/she has been scolded'

b. Ha dato una sgridata a Gianni

Has.AUX given.PART a reprimand to Gianni

'He/she has scolded Gianni'

(18)   a. The Count gives everyone the creeps.

b. You get the creeps just looking at him.   (Richards, 2001: 189)

The fact that some LVCs do not accept non-agentive subjects (19a), while their corresponding synthetic verb does (19b), also supports the thesis that it must be the LV that does contribute to the selection of the external argument (De Miguel, 2008; Sanromán Vilas, 2009), rather than the nominal element (often morphologically related with the synthetic verb).

(19)   a. {Juan/*El viento} dio un golpe al coche

{Juan/*The wind} gave a blow to the car

'{Juan/*The wind} hit the car'

b. El fuerte viento del Cáucaso golpeó el coche en la carretera de Ljubljana

y arrancó el limpiaparabrisas.   (De Miguel, 2008: 571)

'The strong wind from the Caucasus hit the car on the Ljubljana road and

ripped off the windscreen wiper.'

Although this proves that the LV does pose some selectional restrictions (Grimshaw & Mester, 1988), the inability to select a non-agentive subject in (19) does not necessarily depend on the verb.

In fact, the restrictions on the subject with LVCs, as exemplified in (20), must come from the NVE which is proven by the fact that they would not be acceptable either with the synthetic verb, as in (21). This is in line with Marantz's proposal (1984) that the

external argument selection is performed by the verbal predicate (vP), and not only the verb.

(20)   a. {En Joan/*L'oreneta} ha fet una caminada de dues hores

{The Joan/ *The swallow} has.AUX done.PART a walk of two hours

'{Joan/*The swallow} has taken a two-hour walk'

b. {En Joan / *L'oreneta} ha donat un bon consell a la Maria en aquest afer

{The Joan/ *The swallow} has.AUX given.PART a good advice to the Maria on this issue

'{Joan/*The swallow} has given Maria good advice on this issue'

c. {En Joan / *L'oreneta} ha tingut la seva proposta en consideració

{The Joan/ *The swallow}has.AUX had.PART the his proposal in consideration

'{Joan/*The swallow} has taken his proposal into consideration'

(21)   a. {En Joan/*L'oreneta} ha caminat durant dues hores

{The Joan/ *The swallow} has.AUX walked.PART during two hours

'{Joan/ *The swallow} has walked for two hours'

b. {En Joan / *L'oreneta} ha aconsellat bé la Maria en aquest afer

{The Joan/ *The swallow} has.AUX advised.PART well the Maria on this issue

'{Joan/ *The swallow} has advised well Maria on this issue'

c. {En Joan / *L'oreneta} ha considerat la seva proposta

{The Joan/ *The swallow} has.AUX considered.PART the his proposal

'{Joan/ *The swallow} has considered his proposal' (Colominas, 2001: 3)

From a lexicalist perspective, the argument selection in LVCs is explained through an argument transference between the noun and the LV (with origin in the noun). For Sanromán Vilas (2011: 258), the deverbal nominal and the verb it derives from have the same number of semantic actants[6] which can be expressed differently depending on whether they are realised as noun or verb in the surface. For instance, *orden* 'order' in (22a), when it is realised in the LVC *dar una orden* 'give an order', the first semantic

---

[6] Within Explanatory and Combinational Lexicology (Mel'čuk, 1998), semantic actants are the expressions of the participants involved in a lexical unit. For instance, to define *install* there are three semantic actants involved: X, a person, installs Y, a program, on W, a permanent storage device (i).
(i) INSTALL2: X installs Y on Z   (L'Homme, 2007: 189)

actant of the NVE is always the subject, which corresponds to the subject of the synthetic verb in (22b).

(22)   a. La orden de X a Y de Z

'The order from X to Y of Z'

b. X ordena a Y Z

'X orders Z to Y'

The obligatory coreference between the possessive of the NVE in (23a) and the subject of the LVC *Juan* proves that the external argument is semantically related to the NVE itself (Alonso-Ramos, 2004, for Spanish, but also Gross, 1976 and Giry-Schneider, 1987, for French), as well as the ungrammaticality in (23b).

(23)   a. Juan ha dado su/*mi/*tu autorización.   (Alonso-Ramos, 2004: 19)

'Juan has give his/*my/*your authorization'

b. *Monica farà una telefonata di Paolo a Eva. (Alba-Salas, 2004: 288)

Monica will.do a call   of Paolo to Eva

'Monica will give Paolo's call to Eva'

For Sanromán Vilas (2917), there is semantic compatibility between the LV and the NVE in Spanish LVCs by means of lexical features, which are abstract semantic components that are shared between the full and light use of the verb, and between the LV and the eventive NVE (based on the *rasgos léxicos* in Bosque, 2004). The so-called lexical features are elements of semantic compatibility. For example, the LV *dar* 'give' is a verb of transfer which combines with a moved object, as in (24), where all LVCs involve a noun that is transferred or moved from X (agent) to Y (receiver): the *opinion* of the doctor (X) to us (Y), in (24a), and the *advice* of a first person singular to a second person singular, in (24b).

(24)   a. Esperemos a que llegué el doctor, él nos dará su opinión

'Let's wait for the doctor to arrive, he will give us his opinion'

b. Déjame darte un consejo. Aprovecha lo que tienes ahorita y no te metas

en nada raro   (Sanromán Vilas, 2017: 35)

'Let me give you a (piece of) advice. Enjoy what you have right now and

don't get into anything weird'

In fact, there seems to be an underlying path, which is traversed towards a goal. For Sánchez Rufat (2015: 211), both path and goal are present in all uses of the verb *dar* with two different contexts:

1. the objects are pre-existent and are moved (literally or figuratively) until the end both in constructions with *dar* as a full verb (*dar un caramelo,* 'give a sweet') or as a LV (*dar una opinión,* 'give an opinion').

2. the nouns denote an event which is being created and, thus, cannot be transferred or moved, like *beso* ('kiss')*, salto* ('jump') or *clase* ('lesson').

In both cases there is a path involved, but the transfer is only present in the first case when the object is pre-existent. According to Sánchez Rufat (2015), this can also explain the differences between *hacer/dar un masaje* ('make/give a massage'): in *dar* 'give' there is a focus on the recipient, while with *hacer* 'make' the focus is on the action[7].

However, the hypothesis of *argument transference or sharing* is strictly limited to the semantics, and it does not hold in a neo-constructionist approach to argument structure. Myler (2016: 280) defends that the core arguments of event nominals are never transferred in LVCs. His proposal is compatible with that of Choi & Wechsler (2002) for Asian languages (especially Korean and Japanese), where they argue that in (25) the agent (*John*) is assigned a Theta-role by the LV, while the theme argument (*English*) is introduced inside the nominal (between brackets). This goes against what Grimshaw & Mester (1988) had proposed for Japanese, who defended that the LVs inherit the arguments from the nominals.

---

[7] The fact that both LVs are so similar and that the reasons for the choice of one over the other is explained by the emphasis put on certain context could also explain the differences of choice between these two verbs in different languages. Acedo-Matellán & Pineda (2019: 187) highlight that Spanish has a preference over *dar* ('give'), (i), while other Romance languages such as French and Catalan prefer *faire/fer* ('do/make'), (ii).

     (i) *dar miedo* (lit. give fear = to scare)
     (ii) *fer por,faire peur* (lit. make fear = to scare)

Their semantic proximity can also explain the divergences between the norm and the use of LVs *donar* 'give' and *fer* 'make' in present-day Catalan. According to GIEC (2016, 17.2.2), the main differences between the use of the LV *fer* and the LV *give* is related to the agentivity of the subject: with *fer* 'make', the action is created, generated or produced and can be controlled but not necessarily by the subject (as in *fer un petó* 'make a kiss'*, fer broma* 'make joke'); whereas LVCs with *donar* 'give' denote a transfer which is generally controlled by the subject. As pointed out in Alvarez-Morera (2023), the norm focuses on describing the contexts in which the use of *donar* 'give' is incorrect and the verb *fer* 'make' should be used instead, which suggests that there is certain hesitation in their use by Catalan speakers.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

(25)   John-i [yenge-lul kongpu-lul] ha-yess-ta.

          John-NOM English-ACC study-ACC do-PAST-DEC (Korean)

          'John studied English.'

Another piece of evidence proposed by Choi & Wechsler (2002) is that the Korean LV *ha* 'do' selects a complement that must be eventive, and the external argument must be agentive. In this syntactic model, following Kratzer (1996), the introduction of the subject is done via the Voice head above the LV, so the agentive semantics cannot come directly from the nominal. Again, this is in line with Marantz's proposal (1984) that the external argument must be selected by the whole predicate (vP).

In sum, it seems that some of the main contributions attributed to the LV, such as its influence on the aspect and telicity of the construction or some constraints on argument and thematic roles posed by the verb can be attributed to the whole construction instead.

2.2. Properties of the non-verbal element

The internal properties of the NVE are independent of those of the LV, and the differences of use of the LV in light and full uses should be explained through the syntax-semantics of the noun (Colominas, 2001: 39). In this section, some of the properties traditionally attributed to the LV will be discussed as properties dependent on the NVE.

The discussion focuses on the properties of the NVE and how they influence and affect the whole construction. The first aspect is the eventive nature of the NVE and its predicative force. Second, the argument properties of the NVE will be discussed and proven through tests. Finally, a proposal for the classification of the NVE will be put forward.

First, LVs combine with eventive nouns in LVCs (Kearns, 1988; De Miguel, 2008, 2011). These nouns bear their own argument and event structure, but do not assign case to their arguments (Ramchand, 2014). The NVE has been traditionally defined as deverbal (as in 26a), but it can also be a non-deverbal eventive noun[8], at least in Romance

---

[8] Resnik (2010: 297) establishes the following cues for detecting non-deverbal events in Spanish: nouns occuring in PPs headed by *durante* 'during' (i),as arguments of verbs such as *celebrar* 'to celebrate', and (iii) in presence of temporal quantifying expressions such as *dos semanas de* 'two weeks of' (iii), among others.
      (i) durante la clase ('during the lesson')
      (ii) se celebró una actividad ('an activity was celebrated')
      (iii) dos semanas de clases ('two weeks of lessons')

languages (as in 26b). There are also some LVCs with NVE with a resultative reading (as in 26c).

(26)   a. hacer una visita ('make a visit')

         b. dar una clase, hacer una actividad ('give a lesson, do an activity')

         c. hacer una traducción, tomar una fotografía ('make a translation, take a picture')

In the cases where the noun can have a resultative interpretation, they inherit the predicate force of the verb from which they derive (Alonso-Ramos, 2004: 141). For instance, *una traducción* 'a translation' in (27) is a noun whose interpretation is resultative, but it is also a deverbal noun inheriting the event from its base verb (De Miguel, 2011).

(27)   María hizo una traducción de la novela.
         'María translated the novel'

Even without the LV, the NVE keeps their predicative force (as pointed out by De Miguel, 2011, for Spanish LVCs). This is seen in the contrast between the noun phrases with *paseo* 'walk' (28a) and *análisis* 'analysis' (28b), which can select PPs as *por el parque* 'around the park' and *de Luis* 'of Luis', while a referential noun such as *caramelo* 'candy' (28c) cannot select a PP on its own.

(28)   a. El paseo de Luis por el parque
         'Luis' walk around the park'
         b. El interesante análisis de Luis
         'Luis' interesting analysis'
         c. *El caramelo de Luis a su sobrino
         'Luis' candy to his nephew'

The common characteristic of all NVEs in LVCs seems to be, then, that they must have argument structure, and they must have at least one argument to share with the LV (Alonso-Ramos, 2004: 142).

However, this test cannot be definitive, as in (29), where the referential noun *cuadro* 'painting' (29a) can also appear with PP complements, whereas the NVE *orden* 'order' (29b) cannot.

(29)   a. el cuadro de Las Meninas de Velázquez

           'Las Meninas painting by Velázquez'

           b. #el orden de Juan en sus papeles

           'Juan's order in his papers'

According to Cremades (2016: 202-24), there are certain referential nouns in Catalan which accept complements similar to arguments (quasi-arguments), while others reject them. The main difference explaining this divergence in the acceptance or rejection of arguments is connected to the semantic function of theme: when the result of the verbal action is the theme of the predicate, the resultative nominalization does not accept any other argument as theme, as in (31). In contrast, when the noun is not equivalent to the theme of the verbal action, the resultative noun accepts a theme as an argument: in (31a) the analysis (*anàlisi*) is the result of the verb to analyse, but it is not what is analysed (*sang*, 'blood'); likewise in (31b), where the modifications are not what is modified but rather the work (*treball*) is modified.

(30)   a. Hem aprofitat bé aquelles donacions (*de quadres).

           'We have made the most of those donations (*of paintings)'

           b. Les troballes (*de joies) de les tombes del faraons són precioses

           'The findings (*of jewels) in pharaohs' tombs are precious'

(31)   a. Segons aquesta anàlisi de sang tens el colesterol alt.

           'According to this blood test, you have high cholesterol'

           b. Una de les modificacions del treball no m'acaba de convèncer.

           'One of the modifications of the work does not convince me'

Cremades (2014, 2016) suggests that the possibility of resultative nouns in Catalan to admit arguments is related to its equivalence with the theme of the predicate, rather than the aspectual class of the verbal base.

Regarding the nature of nouns in LVCs, Alonso-Ramos (2004: 155-156) argues that it is not necessary for the NVE to be eventive, but it needs to be predicative and have at least one argument. For this reason, the contrast between nouns with argument structure and those without it needs to be made clear.

Grimshaw (1990) established a three-way distinction for the eventivity of the nouns: (i) complex event nominals which denote events with argument structure (*examination*), (ii) simple event nominals which denote events but do not have argument

structure (*race, trip*), and (iii) result nominals which never have argument structure nor an eventive reading (*exam*). However, this distinction is not always straightforward, since some deverbal nouns can have both an event interpretation (32a) and a resultative reading (32b).

(32)   a. the examination of the students by the teacher

         b. a difficult exam(ination) (*of the students by the teacher)

Other classifications have been proposed. While Sleeman & Brito (2010) propose five types of deverbal nominalizations, Borer (2013) focuses on the distinction between argument-taking nouns and result nouns without argument structure. In this line, Alexiadou & Borer (2020: 11) adapt Grimshaw's diagnostics (1990) and classify nouns in two groups: result nominals (both result nominals and simple events) and argument structure nouns (Table 3).

RNs (result nominals & simple events)	ASNs (with argument structure)
No obligatory arguments	Obligatory events (where relevant)
No necessary event ready	Event reading
No agent-oriented modifiers	Agent-oriented modifiers
Subjects are possessives	Subjects are arguments
By-phrases non- arguments (in Spanish *de*)	*By*-phrases are arguments (in Spanish *por*)
No event control	Event control (implicit argument control)
No aspectual modifiers	Aspectual modifiers (for/in 3 hours)
Modifiers (constant, frequent) only with plural	Modifiers (frequent) may occur only with plural
Post-nominal genitive possible	Not possible
Can pluralize & one, a, that determiners	Do not pluralize & do not allow one, that, a determiners
May be predicates	May not be predicates

Table 3. Grimshaw's diagnostics, adapted in Alexiadou & Borer (2020: 11)

Some of the issues that this classification raises affect the analysis of LVCs because simple event nominals can be found in LVCs (*trip* or *race*), and they denote events but according to this framework they lack argument structure. If this is the case, having an argument does not seem to be a necessary condition to be part of an LVC. However, these nouns can appear with complements similar to arguments, both in LVCs (33) and when they are on their own (34).

(33)   a. Our class took a trip to the zoo.

                b. We did a race in Bologna, Italy.

(34)   a. The trip to the zoo was fun.

                b. A race in Bologna was his dream.

Also, Alexiadou (2020) detects that some of the diagnostics in Table 3 are problematic and do not hold cross-linguistically. This is the case for pluralization, as only atelic predicates resist pluralization, while telic predicates accept it, as in (35).

(35)   I heard of repeated killings of unarmed civilians   (Alexiadou, 2020)

Roodenburg (2010) suggests that there are cross-linguistic differences with respect to the acceptance or rejection of pluralization by ASNs. Romance languages like French and Italian do not have the same restrictions as those found in Germanic languages and they do not show an asymmetry between singular and plural argument nominalizations.

In line with Roodenburg's proposal for Romance languages, Melloni (2011) also supports that the pluralization of ASNs in Germanic languages is also accepted, like in the selection of arguments such as *of Cartage* by the plural *destructions* in (36a) and *des Dirigenten* 'of the conductor' by the plural *Umdispositionen* 'arrangements' in (36b).

(36)   a. Each of the three destructions of Cartage (began with a siege)

                b. Die Umdispositionen des Dirigenten zogen sich über Tagen hin.

                'The rearrangements of the conductor went on for days'

                                            (Melloni, 2011: 27)

The possibility to pluralize of ASNs is, therefore, present in both Romance and Germanic language; although pluralization of complex event nominals which keep their arguments are more common in Romance than Germanic languages. In Italian, some nouns can be interpreted both as referential or as eventive: in the result reading (37a), the *traduzione* 'translation' refers to the product of the translating event, as the predicate indicates: *piena di errori gramaticali* 'full of grammatical mistakes'; while in its eventive reading (37b), it can take arguments (*da parte del filologo* 'by the philologist').

(37)   a. La (sua) traduzione (di questo testo) è piena di errori grammaticali.

                'The (/his/her) translation (of this text) is full of grammatical mistakes'.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

b. La traduzione di questo testo (da parte del filologo) è stata lenta e complessa.

'The translation of this text (by the philologist) was slow and complicated'.

(Melloni, 2011: 198-199)

When dealing with the plural of the same noun (*traduzioni*) the same possibilities emerge: in (38a), the noun can only refer to the products of the events of translating, while in (38b) the reading is that of several events taking place (a set of repeated translations), whereas in (38c) both readings could emerge.

(38)   a. Le traduzioni (degli antichi studiosi) sono spesso colme di errori ortografici.

'The (/Ancient scholars') translations are often full of orthographic mistakes'.

b. Le traduzioni della Bibbia (da parte di filologi e traduttori) si sono ripetute nei secoli.

'The Bible was translated repeatedly (by philologists and translators) over the Centuries'.

c. Ripetute traduzioni possono alterare il significato originale dei testi.

'Repeated translations can alter the original meaning of the texts'.

(Melloni, 2011: 199)

In fact, the same readings for Italian can be applied to the Catalan and Spanish equivalent constructions of (38).

Taking into account that some of the diagnostics proposed by Grimshaw (1990) reproduced in Table 3 do not hold cross-linguistically, a more reduced set of tests can be used to determine which one emerges in each context. As proposed by Alexiadou & Borer (2020), Table 4 includes the main diagnostics to identify ASNs: obligatory arguments, compatibility with aspectual PPs such as *in 3 hours* or with *constant/frequent* in singular, as well as the selection of a *by*-phrase as an argument.

Referential nominals	AS- nominals (argument structure)
No event reading	Event reading
Not obligatory	Obligatory arguments
Not compatible with aspectual	Compatible with aspectual "in 3 hours"
Constant/frequent with plural	Constant/frequent with singular
By-phrase is not argument	*By*-phrase is argument
<the form, the exam>	<the examination>
The examination was 10 pages long (referential reading)	The examination of the patient took over one hour (eventive reading)

Table 4. Tests to distinguish referential and AS-nominals (adapted from Alexiadou & Borer, 2020)

Some nominals which appear in LVCs, as in (39), do not pass these key tests when they are on their own: some are not compatible with aspectual modifiers (40), and the expression of an external argument by means of a *by*-phrases is not always possible (41).

(39)   a. make a description, copy, photograph, etc.

          b. fer una traducció, fotografia, disseny, producte, dibuix, troballa, etc.

(40)   a. *Jack's trip for / in an hour

          b. La traducció de Shakespeare {durant / *en} una hora

          'Shakespeare's translation {for/ *in} an hour'

(41)   a. The world trip {*by / of} Jack

          b. La traducció de Shakespeare {*per / de} Joan G.

          'Shakespeare's translation {*by/ of} Joan G.'

However, the eventivity arises when they are part of the LVC, for both aspectual modifiers (42) and *by*-phrases (43), which is due to the LV that incorporates an event reading.

(42)   a. We made a trip for 5 days to Bangkok / We made a trip in 5 hours.

          b. Va fer la traducció del volum {durant/ en} un mes.

          'He/she made the translation of the volume {for/ in} a month'

(43)   a. One trip was made by members of the Commission

          b. La traducció d'aquest volum va ser feta per Joan G.

          'The translation of this volume was done by Joan G.'

That is, the nominal *per se* does not have to accept arguments on its own, because the presence of the LV facilitates the selection of such arguments.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

In fact, LVs have been considered to work as verbalizers, especially in Indo-Aryan languages like Hindi, where the set of synthetic verbs is relatively small compared to the number of LVCs in use (Butt, 2010; Vaidya et al., 2019). It is argued that Hindi speakers are more used to LVCs, since the majority of LVCs in Hindi do not have a synthetic counterpart, they are the most common verbalizers. The use of an LV is also quite common to introduce loan words in a language, which is also the case for certain English loan words in Spanish and Catalan which are combined with LVs, as in (44), while some of them have also developed their own synthetic verb by using the suffix *-ear/-ejar* to create activities, as in (45)[9].

(44)   a. fer un like, fer/donar un zasca, fer spoilers, fer match, fer cringe, etc.

'to like, to slap (verbally), to spoil, to match, to cringe'

b. dar hate, dar zascas, tener swag, hacer salseo, hacer ghosting, etc.

'to hate, to slap (verbally), to have swag, to gossip, to ghost'

(45)   a. likejar, spoilejar, posturejar, etc.

'to like, to spoil, to pose'

b. hatear, salsear, ghostear, etc.

'to hate, to gossip, to ghost'

The use of LVs as verbalizers can also be found in cases where the nominal is clearly referential, but its combination with an LV forces an event reading. For instance, in (46a), the use of the most common LV in Catalan, *fer* 'make', is used with non-event nouns to express activities that are "institutionalized", while in (46b) there is an example of an idiolectal choice of LV *fer* and the noun *Internet* which is used to express an activity: *to surf the Internet*.

(46)   a. fer un cafè, fer una birra

'do a coffee', 'do a beer' (meaning to meet for a coffee and a beer)

b. La meva iaia fa Internet

'My granma does Internet' (meaning to surf the Internet)

In British English, a similar trend can be found with certain non-eventive nouns in combination with the LV *have*. For instance, the adjectives *cheeky* (47a) and *quick*

---

[9] All these are instances often found in social media which have not been accepted in the standard and, to my knowledge, have not yet been studied from a theoretical point of view.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

(47b) can be interpreted as an event modifier rather than a nominal modifier, while the nominal is not of eventive nature in either case.

(47)   a. I had a cheeky ice-cream.   (Twitter)

        b. Let's have a quick cup of coffee.   (Twitter)

In these cases, it seems that it is the presence of the LV which enables the event reading of the nominal and, therefore, contributing to the event structure of the whole construction.

2.3. Interaction between the two elements: LVs and NVEs

As introduced in the previous sections, there are divergences regarding the actual contribution of each component to the full LVC. In this section, the discussion will cover the following aspects: the predication function, the syntactic peculiarities, and the degree of cohesion between the two elements involved in LVCs, the LVs and the NVEs.

2.3.1. The predication function

The usual functions performed by the verb and the noun are altered in the LVC (De Miguel, 2008): while the syntactic head of the LVC is the light verb, the semantic head is the noun. The predication function appears to be shared between both elements of the construction. How this is carried out, however, is explained in different terms depending on the theoretical framework.

De Miguel (2008) puts forward a theory of lexical compatibility between the noun and the LV for Spanish LVCs. This obligatorily presupposes that the verb is not completely devoid of meaning, since it must keep some lexical features that will make it compatible and able to agree with the lexical features of the noun. The *refill hypothesis* from De Miguel (2008: 574) suggests that light verbs are lexically underspecified, and they become specified in combination with the information from the nominal element. That is why the LV is flexible enough to appear in combination with different nouns, but there are some restrictions regarding their co-appearance: they both need to predicate a compatible event. For instance, the aspect in *tener* 'have' is compatibe with permanent states, as in (48a), while *coger 'catch'* with transitory states, as in (48b).

(48)   a. tener odio ('have hate', 'to hate')

        b. coger miedo ('catch fear', 'to get scared')

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

Building on de Miguel's work, Sanromán Vilas (2012, 2013, 2014, 2017) defends the *semantic compatibility* hypothesis. The point of departure is that LVs have lexical meaning, although it is considered to be abstract, and this meaning can be described through its lexical features. According to Sanromán Vilas' (2017: 34) hypothesis, these lexical features constitute the meaning that the LV shares with the full verb and with the predicative nouns with which it cooccurs. That is, the lexical features function as elements of semantic agreement between the nominal element and the LV. For instance, the Spanish LV *dar* ('give') has the same lexical features in its use as a LV (49a) and as a full verb (49b): in both cases it is a verb of transfer combined with a transferred object. In (49a), the information is an abstract entity known by the mum (*la mamá*), which is transferred from her to the dad (*al papá*); likewise, in (49b), the book (*el libro*) is a physical object being moved from the place where the agent is (*la mamá*) to the place where the receiver is (*Aldo*).

(49)   a. Fue   la   mamá quien dio   la   información   al   papá.
was.3SG the.SG mum   who gave.3SG the.SG information to.the dad
'It was the mum who gave the information to the dad'
b. La mamá le   dio   [a Aldo] un   libro editado por los jesuitas.
the.SG mum   him.DAT   gave.3SG [to Aldo] the.SG book published by the.PL Jesuits
'The mum gave Aldo a book published by the Jesuits'

Thus, the same lexical features are shared by both uses of the same verb, according to Sanromán Vilas, which would belong to the same lexical entry: in this case, the polysemic verb *dar* 'give'.

In sum, both de Miguel's *refill* hypothesis and Sanromán Vilas' *semantic compatibility* hypothesis assume that there is feature agreement between the LV and the nominal element. However, such semantic compatibility analysis between both elements cannot be easily applied cross-linguistically, since it cannot fully explain why the same noun, e.g. *walk*, selects different LVs in different languages (50), even in languages that are closely related like Spanish (50b) and Catalan (50c).

(50)   a. Paul took a walk in the park.   (English: take)
b. Paul dio un paseo por el parque.   (Spanish: dar, 'give')
c. Paul va fer una passejada pel parc.   (Catalan: fer, 'do/make')

Although the choice of LV is not the same in all three languages, they all share the same semantics. Also, the event and argument structure of the construction is equivalent: the external argument is agentive, while the internal argument *a walk* is selected by the LV in all three cases.

From a semantic perspective, Colominas (2001: 202) analyses the three verbs (*fer*, *prendre*, *donar*, in Catalan for *make, take* and *give*) as dynamic verbs, which share the agentivity of the event that they select. Regarding the differences in LV selection preferences among Romance languages, for Acedo-Matellán & Pineda (2019: 187), the differences between (50a) and (50b) closeness of the verbs *do/make* and *give*, which are interchangeable in the context of LVCs in several languages, such as Catalan and French preferring *fer/faire* 'do' and Spanish and Portuguese preferring *dar* 'give'. In fact, Sanromán Vilas (2014: 214) argues that in the Spanish cases where the same noun can select both *dar* and *hacer* (i.e. *hacer/dar declaración* 'make/give statement') the main difference lies in that *dar*-LVCs prioritizes the transfer of a message, while *hacer*-LVCs emphasizes the creation of the message.

If this semantic compatibility between the noun and the LV implies a selection of the LV by the nominal element, as defended by Sanromán Vilas (2017), this selection should be represented in the syntax. However, the only categories which can take arguments are verbs and prepositions, since they establish relations between entities. However, nouns and adjectives cannot select arguments[10]; although they may require them semantically, it is not possible for them to introduce them in the syntax (since Hale & Keyser's proposal of argument structure, 1993, 2002). Since the semantic compatibility cannot be expressed in syntactic terms, it should be restricted to the domain of lexical semantics, more specifically, semantic selection at sublexical level.

2.3.2. Syntactic particularities of LVCs

Three syntactic properties discussed in previous studies on Spanish LVCs (Alonso-Ramos, 2004; Bustos Plaza, 2005; De Miguel, 2011) will be reviewed here. First, LVCs accept a double analysis with regards to the PP modification of the NVE, as seen in (51). Second, the nominal element can stand as a nominal phrase without the LV, as in

---

[10] It is true that eventive nouns may take some complements, but these are limited to Prepositional Phrases or possessives (for Spanish, according to Picallo, 1999: 381). In fact, the argument structure of nouns is linked to their event structure, since only eventive nouns can admit some sort of argument selection, as discussed by Cremades (2015: 515) for Catalan nominalizations.

(52). Finally, there are restrictions regarding the typology of determiners that can appear in the NVE (53).

(51)   a. Pedro [tiene] [respeto por Ana]

               b. Pedro [tiene respeto] [por Ana]

               'Pedro has respect for Ana'

(52)   El respeto de Pedro por Ana

               'The respect of Pedro towards Ana'

(53)   Pedro da su/Ø/*mi autorización para salir   (Sanromán Vilas, 2015: 180)

               'Pedro gives his /Ø/*my authorization to leave'

With regard to the PP modification of the NVE, in Spanish LVCs there are two possible analyses of the NVE when followed by a PP (Bosque, 2001: 26). In (54a) the PP is an argument of the LVC, whereas in (54b) the PP is an argument of the NVE (Mendívil, 1999; Alonso-Ramos, 2004).

(54)   a. El paseo que Juan dio por la playa.

               the.SG walk.N that Juan gave.3SG about the.SG beach

               'The walk that Juan took along the beach'

               b. El paseo por la playa que dio Juan.

               the.SG walk.N about the.SG beach that gave.3SG Juan

               'The walk along the beach that Juan took'

Likewise, the analysis of the PP modifying the NVE in English LVCs can be analysed both as an argument of the LVC (55a) or of the NVE (55b), although the former is preferred by native speakers.

(55)   a. The walk (that) we took along the beach.

               b. #The walk along the beach (that) we took.

To account for this difference, Mohanan (2017:8) proposes that the structure of the LVC (e.g. *make a claim*) is of dual nature, as in (56). In contrast, the non-light construction (e.g. *disprove a claim*) has only one structure, as in (57)[11].

---

[11] However, the structure with a relative clause complementing the whole LVC does not hold the clefting (b).

    (i) a. That horses fly Pat did not claim _____.

    b. *That horses fly Pat did not make the claim _____.   (Mohanan, 2017: 7)

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

(56)   a. [[make]V [the claim that S]NP]

         b. [[make the claim] PRED [that S] COMPL]

(57)   [[disprove]V [the claim that S]NP]   (Mohanan, 2017: 7-8)

For Italian LVCs, Alba-Salas (2004: 289) differentiates between the *fare una telefonata* type of LVCs (58), and the *fare paura* type of LVCs (59) regarding the analysis of the PP complements. In line with Bosque's (2001) proposal, the PP *a Eva* in (58) can be analysed as either inside the maximal projection headed by the nominal element *una telefonata* (58b) or as directly dependent of the LV *fare* (58c).

(58)   a. Monica ha fatto una telefonata a Eva.   (Alba-Salas, 2004: 290)
         'Monica called Eva'
         b. Monica le farà una telefonata dopo.
         'Monica will give her a call later'
         c. La telefonata a Eva la   farà   Monica dopo.
         'The call to Eva will be given by Monica later'

However, the PP *a Ali* in (59a) lacks a double analysis, since it can only be analysed as directly dependent of the verb *fare* as an experiencer of the verbal predicate (59b).

(59)   a. Mark fa paura a Ali.   (Alba-Salas, 2004: 299)
         'Mark frightens Ali'
         b. Mark gli fa paura.
         Mark him.DAT does fear.
         'Mark frightens him'

However, it is important to bear in mind that the double analysis is not exclusive of LVCs. Bosque (2001) presents cases in Spanish where the nominal element accepts the double possibility of relativization with certain verbs (60), unlike other full verbs which only accept one possible relativization because in (61) the complement of the noun cannot be extracted from the DP.

(60)   a. La campaña que Nancy {promovió / impulsó} contra el tabaco.
         'The campaign that Nancy {promoted / drove} against tobacco'
         b. La campaña contra el tabaco que Nancy {promovió / impulsó}.

'The campaign against tobacco that Nancy {promoted / drove} against tobacco'

(61)   a. La admiración por los intelectuales que Pedro {critica/ analiza}.

'The admiration for intellectuals that Pedro {criticizes / analyses}'

b. *La admiración que Pedro {critica/ analiza} por los intelectuales.

'The admiration that Pedro {criticizes / analyses} for intellectuals'

(Alonso-Ramos, 2004: 219-220)

The second syntactic characteristic of the LVC is the possibility of the nominal element to stand on its own, as in (62).

(62)   a. Pedro tiene respeto por Ana.

'Pedro has respect for Ana'

b. El respeto de Pedro por Ana.

'The respect of Pedro towards Ana'

What is special of this structure is that the resulting DP preserves the semantic load of the full LVC (63a), while this option seems to be impossible with a full verb (63b).

(63)   a. Luis dio un paseo por el parque.

Luis gave.3SG a walk by the park

'Luis took a walk around the park'

a'. El paseo de Luis por el parque.

'Luis' walk around the park'

b. Luis dio un caramelo amarillo a su sobrino.

Luis gave.3SG a candy yellow   to his nephew

'Luis have a yellow candy to his nephew'

b'. *El caramelo amarillo de Luis a su sobrino.

The candy   yellow   of Luis to his nephew

'Luis' yellow candy to his nephew'

However, this is contradicted by De Miguel's (2011: 141) examples, where a full verb can be omitted while the DP keeps its original meaning (64), as well as some LVs which cannot be supressed (65).

(64)   a. Velázquez pintó el cuadro de Las Meninas.

Velázquez painted.3SG the painting of Las Meninas

45

'Velázquez painted Las Meninas painting'

b. El cuadro de Las Meninas de Velázquez.

The painting of Las Meninas by Velázquez

'Las Meninas painting by Velázquez'

(65)   a. Juan puso orden en sus papeles.

Juan put.3SGPAST order in his papers

'Juan arranged his papers'

b. *El orden de Juan en sus papeles.

The order of Juan in his papers

'Juan's order in his papers'

The fact that not all NVEs in LVCs can stand on their own (65), as well as the fact that some nouns not involved in LVCs do (64), signals that this cannot be claimed as a property of LVCs but rather as a particular property of eventive nouns.

A similar pattern can be found with English nouns: some instances preserve the meaning when the LV is removed (66a), while the same verb in a non-light construction cannot be omitted (66b). With full verbs, a similar pattern can be observed: certain nouns preserve the semantics (67a) and other do not (67b).

(66)   a. Laura made a trip to Moscow.

a'. Laura's trip to Moscow.

b. Laura made a lemonade for the family.

b'. ?/* Laura's lemonade for the family.

(67)   a. Shakespeare wrote 154 sonnets.

a'. Shakespeare's 154 sonnets.

b. Ellen gave them a well-grounded opinion.

b'. *Ellen's well-grounded opinion to them.

Once again, a property which had been attributed to LVCs, when analysed contrastively, proves to be a property of the noun rather than of the LV, or of the whole construction.

Regarding determination, there are some restrictions in the use of prenominal possessive determiners (68). Both the NVE and the LV can have their own complements and adjuncts, but the subject is shared (De Miguel, 2011: 143), so a possessive is only

possible when it refers to the subject. This can also be applied to other languages like English, which does not accept a possessive that does not refer to the subject (69).

(68)   Luis dio {#mi/su} paseo por el parque.

         Luis gave.3SG {#my/his} walk about the.SG park

(69)   He took {#my/his} walk in the park.

Bruening (2016: 57) suggests that this is in line with the phenomenon of control into non-finite clauses, as in (70a) where the verb *decide* enforces obligatory control of the subject of the non-finite clause that functions as its complement. In contrast, in (70b) there is no control: the subject of the non-finite clause is determined by the context rather than the verb of the main clause.

(70)   a. Did you decide [to portray yourself/*oneself/*herself as a victim]?

         b. I believe it would be a mistake [to portray yourself/oneself/herself as a victim].

The same phenomenon can also be observed in LVCs: in (71a), the logical arguments of the nominal element *kick* are controlled by the arguments of the light verb *give*, but in a non-LVC context they are determined by the context (71b).

(71)   a. She *gave him a kick* in the teeth. (kicker = she, kickee = him)

         b. I think a kick in the teeth is warranted. (kicker and kickee determined by context)   (Bruening, 2016: 57)

When the LV has an external argument, this controls the logical external argument of the nominal, but this property is not restricted to LVs.

To sum up, these three syntactic properties, which have been attributed to the idiosyncrasy of LVCs, have proved to be not restricted to these constructions. However, what is responsible for two of these peculiarities (i.e. the double analysis and the possibility of the noun to stand on its own) is the noun rather than the whole construction or the verb on its own. As for the restrictions in determination, these are related to the fact that the subject of the LVC must be coreferential with the external argument of the event expressed by the NVE.

The study of the relationship between the LV and the NVE thus suggests that, when studying these LVCs, the focus should be directed towards the properties of the nominal element.

2.3.3. Degree of cohesion

Having stated the event structure of LVCs and the contribution of both the LV in §2.1 and the NVE in §2.2 to the whole structure, it is important to consider the syntactic relationship between these two elements. LVCs have been classified cross-linguistically in a continuum according to the degree of cohesion between the NVE and the LV (Sanromán Vilas, 2015; Acedo-Matellán & Pineda, 2019). In the proposed continuum there are differences related to syntactic properties of certain nouns and LVs, but there can also be geographical and intraspeaker variation, as it has been found in Basque LVCs (Martinez, 2015; Acedo-Matellán & Pineda, 2019: 212).

At one end of the continuum, there are transitive-like LVCs. For instance, in Basque they select the auxiliary *edun*, the subject bears ergative case, and the NVE is a direct object which can be separated from the LV. The analysis proposed by Acedo-Matellán & Pineda (2019: 204) show that some NVEs in Basque behave as regular DOs (72).

(72)   Aduna-k dantza egin du.

Aduna-ERG   dance do.AUX.PRS.ERG3SG.ABS3SG

'Aduna has danced.'

At the other end of the continuum, some LVCs point toward noun incorporation (73a) because it doesn't accept separation from the LV (73b).

(73)   a. Aduna-k ongi hitz egin du. (Acedo-Matellán & Pineda, 2019: 212)

Aduna-ERG well word do.AUX.PRS.ERG3SG.ABS3SG

'Aduna has talked well.'

b. *Aduna-k hitz ongi egin du.

Aduna-ERG   word   well do.AUX.PRS.ERG3SG.ABS3SG

Meanwhile, there is a group of LVCs in Basque which present an ambiguous behaviour between these two ends (Martinez, 2015).

In the case of Romance languages, only French shows examples of determinerless and caseless NVE. As proposed by Acedo-Matellán (2014) for NVE in certain LVCs in French, there is a case-less noun occupying the object position, which does not allow the coexistence with another object (74), but it cannot be pronominalized (75) nor can it undergo nominal quantification (76).

(74)   a. J'ai froid (*les mains).

I=have cold the hands

b. J'ai faim (*le/*du fromage).

I=have hunger the/of.the cheese

(75)   a. *Je l'ai, le froid. / *J'en ai, du froid.

I it=have the cold I=PARTV have of.the cold

b. *Je l'ai, la faim. / *J'en ai, de la faim.

I it=have the hunger I=PARTV have of the hunger

(76)   a. J'ai {très/ *beaucoup de} froid.

I=have very much cold

'I am very cold'

b. J'ai {très/ *beaucoup de} faim.

I=have very much hunger

'I am very hungry'   (Acedo-Matellán, 2014: 30-31)

These tests point towards an analysis of the nominal as an unincorporated root. In contrast, Spanish and Catalan do not have cases of unincorporated roots.

At this point, it is worth mentioning what Kearns (1988) proposed for English LVCs, distinguishing between true light verbs (*give a groan*) and vague action verbs (*make an inspection*), depending on the status of the complement accompanying the LV. While true light verbs combine with defective nominals, which cannot be passivized (77a), relativized (77b) or pronominalised (77c) and are not definite (77d), vague action verbs appear with argumental DPs.

(77)   a. * A groan was given by the man on the right.

b. ??The groan (which) he gave startled me.

c. ??The deceased gave a groan and gave another one after.

d. *Who gave the groan?

That is, a true light verb complement is seen as an overt realisation of the event, as in (77), whereas the complement of a vague action verb patterns like the direct object of an ordinary transitive (78).

(78)   a. An inspection was made some time last week.

b. The offer (which) the finance company made was surprisingly generous.

     c. The Health Department made an inspection on Monday and may make another one before prosecuting.

     d. Who made the inspection?

However, Macfarland (1995: 81) objects to this distinction[12] since true light verbs pattern like vague action verbs and suggests that LVCs are no different from regular transitive verbs since they accept non-restrictive relative clauses (79a) and quantification (79b), they can be pronominalised if the LV is not repeated (79c) and allow wide scope interpretation (79d).

(79)   a. The man on the right gave a loud groan, which chilled me.

        b. The deceased gave a groan; and then another one just after two.

        c. He gave every groan we heard.

        d. There is a man such that the man on my right gave this groan.

In line with Macfarland (1995), Bruening (2016: 52) argues that the NVE in LVCs behaves like any other NP. First, although the NVE is typically indefinite, this is not necessarily the case, as in (80). Second, the NVE can be passivized (81a) and relativized (81b). Finally, the NVE can be modified in ways not available by the synthetic counterpart (82). These properties mirror those of regular NPs.

(80)   a. a. She gave the kind of sigh that is the result of extreme disappointment.

        b. She gave many sighs as they parted.   (Bruening, 2016: 52)

(81)   a. A deep sigh was given as she slowly lay in the sand.

        b. Most of the sighs that she gives are obviously affected.

(82)   a. She gave him a well-earned pat on the back.

        b. *She patted him on the back well-earnedly.   (Bruening, 2016: 53)

An opposite proposal is that of Mendívil (1999: 86) for LVCs in Spanish, who distinguishes between support verbs (*verbos soporte*) without reanalysis (i.e. *hacer una mención*, 'make a mention/reference') and vicarious verbs (*verbos vicarios*) where the NVE is reanalysed (i.e. *hacer mención*, 'make mention/reference'). The reanalysis of the latter is based on the non-referential nature of the NVE which is bare and, thus, it is not

---

[12] Acedo-Matellán and Pineda's (2019, p. 207) findings after comparing LVCs in Basque and Romance also confirm that Kearns' classification is not a useful one.

able to behave as a syntactic argument, since it does not passivize (83a), cannot be interrogated (83b) nor pronominalized (83c).

(83)   a. *(Una) mención de Luis había sido hecha cuando llegué.

        '(A) mention of Luis had been done when I arrived'

        b. ¿Qué has hecho de Pedro? *(Una) mención.

        'What has Pedro made? *(A) mention'

        c. Luis hizo *(una) mención de Pedro y no la puedo olvidar.

                                                    (Mendívil, 1999: 404)

        'Luis made *(a) mention of Pedro and I cannot forget it'

As in Kearns's distinction, the main difference between both types of verbs (or constructions) is placed on the behaviour of the NVE, and its ability to behave like an argument, rather than on the LV itself.

Taking into account the characteristics of the NVE proposed in previous studies, the data of the present corpus-based study from both Romance (Catalan and Spanish) and Germanic languages (English and German) show that there is a difference in the degree of cohesion between the light verb and the NVE, from transitive-like structures with the NVE introduced by a determiner (84) to more fossilized LVCs with bare nominals (85).

(84)   The committee made a decision during the meeting.

(85)   If you've ever made conversation with someone on a plane.

Some differences between Germanic and Romance languages have been found in relation to the presence of the determiner introducing nominals in object position (as presented in Alvarez-Morera, to appear). The majority of English LVCs are transitive-like with determiner phrases as the NVE (86); whereas most LVCs in Spanish can accept the NVE with and without a determiner (87).

(86)   Eng. take *(a) break, give *(a) description, make *(an) impression

(87)   Sp. tomar (una) nota, hacer (una) pausa, dar (una) ayuda

        'take (a) note, make (a) pause, give (a) help'

This can be explained on the basis of the divergences in determination possibilities in both languages, since Spanish has more flexibility with bare nouns in object position than English (Longobardi, 2001). This general distinction of the language is also found in LVCs, and it is key for the analysis of such constructions. As seen in previous studies,

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

it is the presence (or lack) of determination what is most relevant for the syntactic structure in LVCs since it affects the argument properties of the NVE and, hence, its relationship with the LV.

Since one of the goals of this dissertation is to find regularities and differences in their syntactic-semantic structure in order to create a typology of LVCs in the four languages under study, the analysis will be based on the morphosyntactic characteristics of the NVE: both determination and modification (Chapter 3). After the analysis of corpus data, a more fine-grained classification of LVCs at both ends of this continuum will be presented in the second part of this dissertation with a typology of LVCs for Germanic and Romance language families (Chapter 6).

2.4. Interim summary

This chapter provided an overview of the properties of LVCs, as well as the properties of its two different components: LVs and NVE. In this review of the treatment of these constructions, I have noticed that, despite being a recurrent topic, the properties attributed to LVs are different across languages.

The first section has discussed the properties of the LV itself (§2.1). First, the focus has been placed on the identification of the LV. For this purpose, the LV has been compared to other verbal predicates with which it could be confused: auxiliaries, serial verbs and main full verbs. The main differences between LVs and auxiliaries relate to form (since LVs are not defective in their paradigm) and aspectual contribution (as LVs are related to the argument and event structure of the whole construction, not just adding aspectual features like auxiliaries). Likewise, they differ from serial verbs with similar arguments related to form and aspect. Finally, LVs are identical in form to a main full verb, from which it is distinguishable only by its semantic weight. Second, the contribution of the LV to the whole construction has been discussed. After analysing the influence of the LV to the argument and event structure of LVCs across languages, I conclude that some of the main contributions attributed to the LV (i.e. aspect, telicity and some constraints on argument structure such as the selection of the subject) should be attributed to the whole predicate instead.

The second section has focused on the properties of the NVE (§2.2). The most highlighted property of these nominals refers to their eventive nature and predicative force, for which they are mostly deverbal –although not necessarily. Basic argumenthood

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 2. Properties of light verb constructions

tests in the Germanic and Romance languages under study have shown that these NVE are arguments of the LVs. The particularities of these arguments, however, depend on some internal properties of the nominal domain, which are mainly related to determination and modification. These are, therefore, the elements for which the empirical dataset is controlled in the next chapters.

Finally, the third section has discussed the issues that arise in the interaction between both components of the LVC (§2.3). First, the LVCs predication function has been a matter of debate because the regular functions of the verb and the noun are altered in these structures. However, the LV is still the head of the verb phrase and its semantic compatibility with the noun allows that they both share some arguments of the predication. Second, the syntactic particularities attributed to LVCs in previous studies (i.e. the double analysis, the NVE as an independent nominal phrase, and some restrictions regarding the determination) have been shown to be rather due to the eventive nature of the NVE, and other general control properties of the language.

All in all, the chapter has set the foundations for the analysis of the empirical data obtained for this investigation, which will be reported in the next chapters. The conclusions drawn from the quantitative and qualitative analysis of the dataset will provide insights into the issues examined here, i.e. the properties of LVs, NVEs and the LVC as a whole.

Chapter 3. Empirical study

This chapter reports the corpus-based study on light verb constructions in Germanic and Romance languages which constitutes the empirical basis of this investigation. In the upcoming subsections, the results obtained will be elaborated, but also the methodological assumptions which build the present study. To this end, the basic methodological approach is described in §3.1, where a line will be drawn between doing linguistics with a corpus and doing corpus linguistics. Then, the main methodological decisions will be clarified in §3.2., which will revolve around the data constituting the object of study (§3.2.1), the selected corpora (§3.2.2), the data retrieval procedure (§3.2.3) and the data analysis (§3.2.4). In §3.3. I report the data analysis performed in the quantitative results presented in the following subsections: *give*-LVCs (§3.4), *take*-LVCs (§3.5) and *make*-LVCs (§3.6). The final sections of this chapter will provide a statistical analysis of the results for each language (§3.7) and outline the main conclusions that can be drawn with a few additional remarks relevant for theoretical linguistics (§3.8).

3.1. Doing linguistics with a corpus

Corpus linguistics has been defined as the study of language based on real life language use (McEnery & Wilson, 2001: 1) that take the form of corpora, "a collection of texts that has been compiled to represent a particular use of a language and it is made accessible by means of corpus linguistics software that allows the user to search for a variety of language features" (Cheng, 2012: 6). More specifically, a corpus-based approach to linguistics uses corpus data to validate, refute or refine previous linguistic theories (McEnery et al., 2006, McEnery & Hardie, 2012: 6).

With this aim, the present dissertation seeks to analyse a specific construction, LVCs, on the basis of usage-based instances found in corpora. These data are further contrasted against grammatical judgments from native speakers and information found in grammar, dictionaries and previous studies. In addition, relevant syntactic and semantic tests are applied to the instances found in corpora for a more detailed analysis of the properties of LVCs (see Chapter 6).

This methodological approach thus aligns with doing linguistics with a corpus, rather than doing corpus linguistics. According to Bosque (2004), the main difference between both approaches is that corpus linguistics relies on the texts and the statistics

applied to the data, while it rejects the speculative considerations of theoretical proposals (Bosque, 2004). In this same line, Egbert et al. (2020: 51) highlight the importance that "researchers also examine the actual linguistic patterns in the texts themselves to explain and interpret the numeric trends observed." Thus, their proposal is that for quantitative corpus linguistics analysis to be useful, it must be accompanied by a qualitative interpretation of the results (Egbert et al., 2020: 69).

The focal point of this approach to corpus linguistics is, thus, language itself: studying in detail a certain linguistic pattern, which is LVC in this case. The data from corpora and statistical methods are basic to achieve this goal, but additional work is needed. In Egbert et al.'s words:

> Language is, and should remain, the primary focus of corpus linguistic investigations. There is no denying that statistical methods are very useful– and often times necessary–for detecting tendencies that might otherwise go unnoticed. However, sophisticated statistical methods often create layers of distance between corpus researchers and the language data they aim to describe, which is likely to negatively affect the linguistic validity of the results. Put differently, any kind of abstraction away from the language data comes with an opportunity cost because it increases the risk of obtaining linguistically uninterpretable results, which, in turn, is more likely to lead to misinterpretations and unsatisfactory conclusions. (2020: 51)

After all, doing linguistics with a corpus represents a methodological approach where the corpus is at the service of the researcher, that is, the data from corpora must be filtered by the linguist's introspection and complemented with other sources (Bosque, 2004). In fact, the balance is found through the triangulation of linguistic behaviour found in corpora, the general constraints from grammar, and the native speakers' judgements of well-formedness and acceptability, which are all complementary and equally important according to Pullum (2017: 296).

In this chapter, the data extracted from corpora for the four languages under study (English, German, Catalan and Spanish) will be presented, followed by a discussion which contrasts the present results and those found in previous studies on LVCs.

3.2. Data and corpora

This section focuses on the main methodological issues of the case study. The data that constitutes the object of study will be described in detail in §3.2.1, and the corpora from which the dataset was retrieved will be introduced in §3.2.2. Finally, the data

retrieval procedure will be described as well as the methodological decisions taken during this process in §3.2.3.

3.2.1. The object of study

The object of study of this dissertation are LVCs in four different languages: English, German, Catalan and Spanish. More specifically, the selected verbs are *give, take* and *make*, and their corresponding verbs in each language: *geben, nehmen* and *machen*, in German; *dar, tomar* and *hacer*, in Spanish; and *donar, prendre* and *fer*, in Catalan.

As detailed in the introduction of this dissertation (Chapter 1), a crosslinguistic approach for such language combination has not been addressed so far. The contrast between Germanic and Romance languages is particularly relevant due to their divergent behaviour in the selection of (bare) nominals in object position, which is due to differences in their semantic reading and syntactic limitations (Chierchia, 1998; Longobardi, 2001).

For any empirical study, every linguistic feature requires an operational definition (Biber & Conrad, 2019). Since a comprehensive introduction to LVCs has already been put forward in Chapter 2 and the details on the data retrieval will be described in §3.2.3, the remainder of this section will present the LVCs retrieved and analysed for this investigation. Besides, data from previous studies on specific verbs in certain languages will be provided alongside the reasons for the choice of combinations included in the analysis.

Since one of the goals of this dissertation is to find out regularities and differences in their syntactic-semantic structure in order to create a typology of LVCs in the four languages (English, German, Spanish and Catalan), the basis for the proposed classification will be the morphosyntactic characteristics of the NVE in relation to both determination and modification.

A different classification of LVCs also based on the NVE is the one proposed by Live (1973), and followed by Levin & Ström Herold (2015), where three types of LVCs are proposed according to the derivational properties of the noun:

1.   Type I: light verb + deverbal/zero-derived noun: TAKE a walk, GIVE sb. a call, TAKE command

2.   Type II: light verb + suffixed noun: MAKE an apology, GIVE an illustration

3.   Type III: light verb + noun lacking corresponding verb: TAKE a toll, MAKE an effort, GIVE the creeps

Considering this classification of the NVE, Levin & Ström Herold (2015) establish significant differences between three Germanic languages: English, German and Swedish. In their study, English shows a clear predominance of Type I and it reflects the easiness of the language to convert verbs to nouns by zero derivation. In contrast, German and Swedish present a more predominant Type II, which derives from a richer derivational morphology than English. Finally, the main difference between the latter two languages is the higher presence of nouns without an equivalent synthetic verb in Swedish LVCs. However, in their analysis of parallel texts, Levin & Ström Herold (2015: 23) find out that only Swedish translations from and into English seem to be affected by the type of LVC according to the morphology of the noun, while this is not the case for the analysed German LVCs, which differ more from the English LVCs and do not correlate with the predominant types in this language. When they look into the modification of the NVE in the three languages, however, there is no correlation between this and the type of LVC in Live's (1973) classification. Therefore, the morphology of the noun is not taken as a variable in the present corpus-based analysis.

With respect to the choice of the three verbs (*give, take* and *make*, as well as their corresponding verb in each language), this is motivated by their frequency in the four languages under study. According to Butt & Lahiri (2003), the list of the most frequently used verbs is shared crosslinguistically: *come*, *go*, *take*, *give*, *hit*, *throw*, *give*, *rise*, *fall* and *do/make*. From this basic repertoire, the three verbs selected are present in all languages in this empirical study (1).

(1) a. Eng. give a kiss, take a picture, make a claim

b. Ger. eine Antwort geben, die Rücksicht nehmen, Bemerkung machen

'give an answer, take (into) consideration, make a remark'

c. Cat. donar consell, prendre una decisió, fer un petó

'give advice, make/take a decision, give a kiss'

d. Sp. dar un paseo, tomar una determinación, hacer un cambio

'give a walk, make/take a determination, make a change'

Although there is a clear convergence in what constitutes the basic repertoire of LVs crosslinguistically, there are some differences and preferences of certain verbs depending on the language (i.e. *fer/faire* 'make' is preferred in Catalan and French versus *dar* 'give' in Spanish, Acedo-Matellán & Pineda, 2019: 185), or even a variety within the languages (as studies of World Englishes have proven, i.e. Giparaite, 2016; Mehl, 2017).

First, the light verb *give* is quite stable in terms of frequency since early modern English (Butt, 2010; Elenbaas, 2013), and it has been recurrently studied as a LV since Kearns (1988). In contrast, the German verb *geben* is not the most common LV, and it is not found among the most frequent verbs of the language (Bruker, 2011: 45). However, a synonymous verb *erteilen* is placed in the fifth position of the most common LV, which is a more formal alternative of *geben* (Winhart, 2001; Storrer, 2007). Other studies (Levin & Ström Herold, 2015) have chosen the LV *geben* instead of its formal variant *erteilen* for a comparison with the English LV *give* because the presence of *erteilen* is only high in scientific and formal literature corpora. For this reason, the LV *geben* will be examined in the present dissertation. In the Romance languages, the LV *dar* in Spanish and *donar* in Catalan are also among the most common verbs with certain differences. In Spanish, *dar* is the most frequent verb in the language (RAE/ASALE, 2009: §34.7.2; Sánchez Rufat, 2015), and it is also the LV with more possible combinations (De Miguel, 2010). In contrast, *donar* in Catalan is not the most basic LV, although it is also present in a considerable amount of LVCs of the language (GIEC, 2016).

Second, the light verb *take* has also been widely studied as one of the most frequently used verbs in the English language, since Huddleston & Pullum (2002). In German, the LV *nehmen* is one of the most frequent verbs of the language (Bruker, 2011), and it has been studied in several studies as one of the most paradigmatic LVs of the German language (Levin & Ström Herold, 2015; Fleischhauer & Gamerschlag, 2019). In comparison, the light verbs *tomar*, in Spanish, and *prendre*, in Catalan, seem to be not so frequent. Although they are still encountered under the lists of the most common LVs (Sanromán Vilas, 2017; GIEC, 2016: 17.2.2), their possibilities of combination with nominal elements are more restricted than other LVs.

Third, the light verb *make* is also present in the basic crosslinguistic repertoire alongside its almost synonymous light verb *do*, which has been discarded for this study

due to its alternative nature as an auxiliary verb[13]. In present-day English, the LV *make* is quite present (Perea Irigoyen & Sánchez Róo, 2013), although it coexists with the LV *take* in some LVCs (i.e. *make a decision* (American English) vs. *take a decision* (British English), Colominas, 2001), which makes it an interesting verb to study. For German, the verb *machen* has been classified as the most common German LV (Bruker, 2011: 45), and it also coexists with its informal variant *tun*. In this case, the more standard verb *machen* was selected for practicality in searching in the selected corpus for the German language, where the majority of texts are written and, thus, the register is formal. In turn, the Romance variants *hacer* and *fer* are very present in the Spanish and Catalan LVCs (Sanromán Vilas, 2014; GIEC, 2016). It should be noted that there is a difference in the relevance of this verb in these languages, since the verb Catalan LV *fer* is considered to be the default LV in the language due to its productivity (Ginebra & Navarro, 2015; De la Cruz, 2021).

This section has provided an overview of the complexity of the set of LVCs selected for analysis in this dissertation, as well as a brief summary of the diversity of findings and treatment that the different LVs have been given in the literature.

3.2.2. Corpora

The data for this study is based on instances from four representative and extensive electronic 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). All four corpora are available online, open access and syntactically annotated. In this section, I motivate the criteria for the selection of these specific corpora, while describing, assessing, and contrasting their main properties.

There are certain differences between these corpora, mainly based on the number of words and the time span they cover (see Table 5). With the aim to investigate present-day use of LVCs in the four languages and to research a homogeneous sample, the time frame set for the extraction was delimited to the 21st Century.

---

[13] As will be stated in the data retrieval section (3.2.3), the data pruning had to be done manually, due to the incapacity of the used corpora tools to tell LVCs apart from other similar constructions. The reason to discard the LV *do* and, thus, select the LV *make* instead, is of practical reasons: to avoid the retrieval of instances with the auxiliary *do*.

Lang.	Corpus	Size	Genres	Time
English	COCA[14]	+1 billion words of text (25+ million words each year 1990-2019)	8: spoken, fiction, magazines, newspapers, academic texts, blogs, and TV and movies subtitles.	2000-2019
German	DWDS[15]	28 billion words of text (1900-2020)	6 genres: spoken, fiction, newspapers, academic texts, blogs, and TV and movies subtitles.	2000-2020
Catalan	CTILC[16]	+100 million words of text (1832-2018)	3 genres: fiction, newspapers and academic texts.	2000-2018
Spanish	CORPES XXI[17]	312 million words of texts (21st Century)	6 genres: spoken, fiction, newspapers, academic texts, blogs and others.	2000-2020

Table 5. Comparison of the corpora

The representativeness of a corpus is based on its capacity to generalize the data of a language or a variety (Leech, 1991: 27), as well as its variability (Biber, 1993: 243). In order to evaluate their representativeness, a summary of the metadata, documentation and files of each corpus will be presented. However, it should be noted that the representativeness of a corpus should be interpreted in relative terms since there are no objective criteria to balance a corpus (Egbert et al., 2022: 102), nor do all corpora make the same kind of information available.

COCA contains more than one billion words of data covering the period 1990-2019. According to Davies (2008), it is "the only corpus of English that is 1) large 2) recent and 3) has a wide range of genres". The number of words in the corpus is distributed as follows: approximately 20 million words per year and around 4 million

---

[14] Available at: https://www.english-corpora.org/coca/ (Data retrieved 11.10.2021-02.11.2021, and 06.07.2022-23.07.2022)

[15] Available at: https://www.dwds.de/r (Data retrieved 16.11.2021-01.12.2021, and 10.09.2022-23.09.2022)

[16] Available at: https://ctilc.iec.cat/ (Data retrieved 11-27.05.2021, 13-23.12.2021, and 24.09.2022-01.10.2022)

[17] Available at: https://www.rae.es/banco-de-datos/corpes-xxi (Data retrieved 02.11.2020-15.12.2020, and 08.10.2022-11.11.2022)

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study

words per genre. Table 6 shows the composition of the corpus for the time period analysed in this dissertation (2000-2019), to which a general amount of 125,496,215 words from blogs and 129,899,426 words from the web must be added.

YEAR	TV/FILMS	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	TOTAL
2000	4,385,593	4,455,815	3,942,474	4,387,935	4,037,086	4,093,991	25,304,894
2001	4,389,164	4,026,240	3,894,789	4,298,636	4,072,447	3,965,654	24,648,931
2002	4,384,475	4,372,290	3,766,673	4,310,634	4,114,280	4,054,359	25,004,713
2003	4,386,799	4,445,270	4,125,039	4,332,708	4,056,245	4,047,802	25,395,866
2004	4,378,535	4,359,084	4,099,691	4,337,309	4,121,117	4,009,359	25,307,099
2005	4,382,594	4,438,877	4,101,737	4,364,776	4,124,225	3,925,927	25,340,141
2006	4,369,684	4,345,995	4,113,173	4,302,713	4,120,732	4,019,200	25,273,503
2007	4,384,406	3,914,424	4,063,116	4,225,511	4,002,299	4,303,993	24,895,756
2008	4,376,702	3,467,315	4,147,216	4,289,641	4,021,006	3,977,790	24,281,678
2009	4,360,676	3,942,512	4,072,580	3,972,290	3,956,523	3,975,128	24,281,718
2010	4,386,795	4,097,760	3,897,459	3,832,576	4,226,666	3,838,637	24,281,903
2011	4,366,464	4,706,635	4,165,068	4,194,966	3,941,853	4,474,072	25,851,069
2012	4,379,595	4,411,281	3,862,889	4,306,912	4,126,669	4,384,263	25,473,621
2013	4,379,396	3,986,106	4,256,880	4,190,854	4,106,654	3,559,748	24,481,651
2014	4,380,134	3,850,683	4,172,260	4,264,503	4,140,151	3,476,429	24,286,174
2015	4,377,018	3,980,660	4,218,823	4,205,807	4,108,436	3,638,406	24,531,165
2016	4,380,381	4,168,303	3,258,473	4,053,156	4,059,857	3,968,779	23,890,965
2017	4,384,822	4,225,248	3,940,337	4,212,809	4,154,518	4,052,435	24,972,186
2018	4,353,912	4,300,990	4,109,362	4,143,311	4,158,845	4,200,047	25,268,485
2019	4,350,065	4,300,658	4,099,130	4,152,861	4,123,367	4,080,106	25,108,206

Table 6. Composition (in words) of COCA (2000-2019)[18]

Apart from its size and the up-to-date data, other features facilitate its use for a corpus-based study, such as frequency information for certain words or the possibility to search for collocates and concordances, a key aspect of this dissertation. Moreover, as argued in Hundt et al. (2007: 1), there is a lack of representative corpora for varieties of English other than American and British English, for example, or for the recently developed genres (i.e. chatroom discussions, text messaging). This current situation makes it necessary to resort to the limited corpora and material available. In this regard, the corpus selected for this dissertation is one of the most complete and representative of those that are available online with open access.

---

[18] Available at https://www.english-corpora.org/coca/ (Section: Text and genre)

DWDS is an open-access collection of electronic resources of word usage in German with historical and contemporary data. It is regularly updated by the *Berlin-Brandenburgischen Akademie der Wissenschaften* (Berlin-Brandenburg Academy of Sciences). The *Kernkorpus21* (core corpus of the 21st Century) is balanced by time and genre: each decade includes approximately 100,000 words divided among four text types: fiction, nonfiction, newspapers and academic texts. It also contains specialized corpora with film subtitles, blogs and transcribed oral speeches. Table 7 shows the composition of the corpus for the time period analysed in this dissertation (2000-2020).

DECADE	FICTION	NONFICTION	NEWSPAPERS	ACADEMIC	SPOKEN	BLOGS	TV/FILMS	TOTAL
2000	3,443,291	360,987	10,754,643	803,210	10,865,951	38,212,714	2,687,860	67,128,656
2010	33,709	-	-	-	11,514,981	70,307,329	21,468,353	103,324,372

Table 7. Composition (in tokens) of DWDS (2000-2020)[19]

All varieties of German are represented in the corpus, although the data for the exact geographical distribution of each subcorpus is not available. Thus, geographic variation cannot be taken into consideration.

CTILC is an open-access corpus of written language developed by the *Institut d'Estudis Catalans* (Institute of Catalan Studies) with more than 100 million words from 1832 to 2018. The texts are classified in three genres: fiction, newspapers and academic texts. Table 8 shows the composition of the corpus for the period of time analysed in this dissertation (2000-2018), which includes a total of 23,529,322 words which represent approximately 1,000 words per year.

---

[19] Available at https://www.dwds.de/r (Section: Krpora im DWDS)

YEAR	FICTION	NONFICTION (press and academic)	TOTAL
2000	465,297	1,080,933	1,546,230
2001	685,939	941,499	1,627,438
2002	402,299	369,724	772,023
2003	305,176	588,703	893,879
2004	724,546	751,497	1,476,043
2005	643,377	804,506	1,447,883
2006	518,826	564,488	1,083,314
2007	244,180	770,455	1,014,635
2008	355,907	717,588	1,073,495
2009	294,832	756,958	1,051,790
2010	391,386	1,028,822	1,420,208
2011	781,428	707,587	1,489,015
2012	638,647	603,483	1,242,130
2013	460,672	535,846	996,518
2014	967,570	762,318	1,729,888
2015	811,776	1,113,082	1,924,858
2016	231,548	984,312	1,215,860
2017	282,451	743,368	1,025,819
2018	387,111	111,185	498,296

Table 8. Composition (in words) of CTILC (2000-2018)[20]

All varieties of Catalan are present in the corpus, but they are unevenly represented: Central Catalan represents 72.32% (17,015,479 words); Valencian, 11% (2,587,779 words); Northwestern, 8.31% (1,955,987 words); Balearic, 6.88% (1,618,660 words); Northern, 1.01% (238,698 words), and Alguerès, 0.48% (112,719%). Although the number of native speakers of Catalan is also higher in the region of Catalonia (which includes the Central Catalan and Northwestern varieties), the proportion of representation in the CTILC corpus does not correspond to the distribution of native speakers in the different varieties. This could have influenced the overall results of this study, especially when different varieties of the language favour the use of some LVs over others (i.e. in Valencian and Balearic Catalan, the LV *donar* 'give' is present in more LVCs than in Central Catalan).

---

[20] Available at https://ctilc.iec.cat/ (Section: Dades numèriques)

CORPES XXI is a reference corpus of the 21st Century compiled by the *Asociación de Academias de la Lengua Española* (Association of the Spanish Language Academies) with a general distribution of approximately 70% for American texts and 30% for Peninsular Spanish, and it also includes some texts from Equatorial Guinea and the Philippines. The texts are 90% written and 10% oral. The former is divided in 40% books, 40% magazines and periodic publications, 7.5% Internet material and 2.5% others. Table 9 shows the composition of the corpus for the time period analysed in this dissertation (2000-2020), which includes around 312 million words which represent approximately 25 million words per year[21].

YEAR	TOTAL
2001-2005	119,401,364
2006-2010	124,258,076
2011-2015	94,603,897
2016-2020	58,108,160

Table 9. Composition (in words) of CORPES (2000-2020)[22]

The information available for each corpus differ and, thus, a clear-cut comparison cannot be provided. Also, there are some variables which could potentially affect the general comparison of the results. First, Tables 5-9 have shown the quantitative differences among the different corpora. In an attempt to ensure representativeness and balance, more data from the bigger corpora were retrieved (English, COCA and Spanish, CORPES XXI), while results from German DWDS and Catalan CTILC are the lowest. This is controlled by the selection of LVCs lemmas (see §3.2.3.) as well as the limitation of instances per each lemma that are found in the smaller corpora.

Moreover, the registers included in the corpora differ, as well as their classification (see Table 5). The vast majority of the data is, however, from written sources and oral language is very little represented in the four corpora. This limits the representativeness of the empirical data, but it is a general limitation of the corpus linguistics field of studies.

With regard to dialectal variation, the corpora show different tendencies: the two corpora for the Romance languages have chosen to balance the representation of the varieties in relation to the number of speakers for each dialect; while the two corpora for Germanic languages do not present clear numbers of representation: the COCA focuses

---

[21] The version used in this dissertation is 0.92

[22] Available at https://www.rae.es/banco-de-datos/corpes-xxi/parametros-de-seleccion-de-textos

entirely on American English, but it does not provide more specific geographical information, while the DWDs does not give the general proportion of each variety of German but it includes the geographical area for each source.

Although register and geographical variation is relevant for corpus linguistics, it has not been taken as a variable for the quantitative analysis, which has focused mainly on grammatical variables (see §3.2.3). It has been taken into account when relevant during the qualitative discussion of the results, as well as the typology proposal in the second section of this dissertation.

Finally, it is important to note that the four corpora have an institution behind which is responsible for the representativeness of the language found in the corpus: *Berlin-Brandenburgischen Akademie der Wissenschaften* (BBAW) for German, *Institut d'Estudis Catalans* (IEC) for Catalan, *Real Academia Española* (RAE) for Spanish, as well as Mark Davies, a Professor of Linguistics. These institutions select the parameters and the codification of the specifically designed corpus, and they aim to ensure that the corpus represents the general characteristics of the languages at a certain moment in History, namely the first decades of the 21st Century. All in all, these corpora have also been used in several previous corpus-based studies, as well as theoretical linguistics investigations based on everyday language use.

3.2.3. Data retrieval

The object of the study is 153 LVCs in four languages, as in Table 10. The final selection of individual LVCs to be analysed for each language was set to a maximum of 15 LVCs per language, and each LVC was classified according to its frequency in the corpus. The total for each LV is in a similar range: 50-52 individual LVCs; while the distribution among languages is around 40, with the exception of German (n=29), which is due to the prevalence of LVCs with a NVE introduced by a preposition in this language (not included in the present dissertation).

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study

	English	German	Catalan	Spanish	TOTAL LV
Give | Geben | Donar | Dar	13	9	14	15	51
Take | Nehmen | Prendre |Tomar	15	8	14	13	50
Make | Machen | Fer | Hacer	14	12	13	13	52
TOTAL LANG	42	29	41	41	153

Table 10. Distribution of the number of lemmas of the database for each language

The final selection of *give*-LVCs to be analysed for each language is represented in Table 11. The highest number of lemmas (n=15) is found in Spanish *dar* LVCs, whereas Catalan (n=14), English (n=13) and German LVCs are slightly fewer (n=9). This is due to the fact that LVCs are more diverse in Spanish than in other languages (Rácz et al., 2014).

	High frequency	Medium frequency	Low frequency
Give (English)	advice, speech, answer, example	try, hug, smile, kiss	notice, nod, push, chase, sigh
Geben (German)	Hinweis ('indication'), 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')

Table 11. *give*-LVCs analysed in the four languages under study

The final selection of *take*-LVCs is represented in Table 12. The highest number of lemmas (n=15) is found in English *take*-LVCs followed by Catalan (n=14) and Spanish (n=13), while German presents the fewer lemmas (n=8). The majority of *nehmen*-LVCs in German have a prepositional phrase involved, which were discarded due to the different behaviour of the nominal within a prepositional phrase and a nominal phrase (i.e. *in Besitz nehmen* 'take possession', *in Empfang nehmen* 'to take receipt/to receive', *in Gebrauch nehmen* 'to make use', *in Haft nehmen* 'to take into custody', *zur Kentniss nehmen* 'to take note', *in Schutz nehmen* 'to take protection').

	High frequency	Medium frequency	Low frequency
Take (English)	step, advantage, action, picture, breath	risk, walk, shower, bite, nap, bath	hike, drink, stroll, swim
Nehmen (German)	Abschied ('farewell'), Rücksicht ('consideration'), Einfluss ('influence'), Platz ('place/seat'), Abstand ('distance')	Einsicht ('insight'), Anteil ('interest')	Anstoß ('ofence')
Prendre (Catalan)	decisió ('decision'), consciència ('consciousness'), nota ('note')	possessió ('possession'), iniciativa ('iniciative'), partit ('party'), precaució ('precaution'), molèstia ('bother')	distància ('distance'), determinació ('determination'), bany ('bath'), resolució ('resolution'), impuls ('boost'), represàlia ('reprisal')
Tomar (Spanish)	decisión ('decision'), nota ('note'), fotografía ('picture'), conciencia ('consciousness')	iniciativa ('iniciative'), precaución ('precaution'), distancia ('distance')	determinación ('determination'), descanso ('rest'), baño ('bath'), acuerdo ('agreement'), represalia ('reprisal'), atajo ('shortcut')

Table 12. *take*-LVCs analysed in the four languages under study

The final selection of *make*-LVCs is represented in Table 13. The highest number of lemmas (n=14) is found in English *make*-LVCs followed by Catalan (n=13) and Spanish (n=13), while German presents the fewer lemmas (n=12).

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study

	High frequency	Medium frequency	Low frequency
Make (English)	Decision, mistake, choice, call, progress	Claim, contribution, profit, payment	Adjustment, announcement, assumption, accusation, mention
Machen (German)	Gedanke ('thought'), Sorge ('worry'), Fehler ('mistake'), Angabe ('detail')	Erfahrung ('experience'), Hoffnung ('hope'), Anfang ('beginning')	Foto ('picture'), Sport ('sport'), Ausführung ('implementation'), Beobachtung ('observation'), Eindruck ('impression')
Fer (Catalan)	Referència ('reference'), esforç ('effort'), feina ('work/job')	Ús ('use'), pregunta ('question'), visita ('visit'), viatge ('trip'), petó ('kiss')	Anàlisi ('analysis'), dibuix ('drawing'), fotografia ('picture'), abraçada ('hug'), aclariment ('clarification')
Hacer (Spanish)	Pregunta ('question'), esfuerzo ('effort'), referencia ('reference') uso ('use')	Viaje ('trip'), ejercicio ('exercise'), análisis ('analysis') declaración ('statement')	Carrera ('run'), énfasis ('emphasis'), crítica ('review'), broma ('joke'), colección ('collection')

Table 13. *make*-LVCs analysed in the four languages under study

Given the lack of largely annotated corpora which enable the automatic retrieval of LVCs, it was necessary to use the search system for collocates in every corpus. The sample was obtained by searching for the collocation of LVs under analysis and the specific nominal elements which build each LVC[23]. The following tools available in each corpus were used to extract the collocates: *Collocates* page in COCA, *DiaCollo* section in DWDS, *Col·locacions* in CTILC, and *Coaparición* in CORPES XXI.

---

[23] As stated in the Introduction (Chapter 1) of this dissertation, the empirical basis of this study are LVCs with an eventive reading. Therefore, LVCs with a causative reading are discarded as they have already been studied within causative constructions (Alsina, 1996; Alba-Salas, 2004).

The instances retrieved were exported to an Excel spreadsheet to be coded for a set of variables based on the relevant literature[24]. The examples of non-light structures were discarded by manual pruning. For example, cases like those in (2) were removed because the nominal element is combined with a non-finite form of the LV, while those in (3) are instances where a preposition is included in the LVCs[25].

(2) a. *Los viajes hechos* entre 1969 y 1972 probaron que un nuevo mundo pudo ser alcanzado. (CORPES XXI, 2019)

'The trips made between 1969 and 1972 proved that a new world could be reached.'

b. le abonó el terreno a *la decisión tomada* poco tiempo después por la madre (CORPES XXI, 2001)

'(he/she) laid the foundation for the decision taken a bit later by the mother'

(3) a. El hombre se dio cuenta y *me dio de besitos* toda la noche.

(CORPES XXI, 2006)

'The man understood and gave me (of) little kisses all night long'

b. te hizo señalarlos y describirlos hasta que empezó a *darte de bofetadas*

(CORPES XXI, 2004)

'(s/he) made you point and describe them until (s/he) started giving you (of) slaps'

c. —¿Si ya no lo fuera, qué harías, *darme de puñaladas*? (CORPES XXI, 2001)

'If I were not anymore, what would you do, give me (of) stabs?'

After manually excluding instances of non-light structures, for every LVC a sample of up to 200 occurrences were analysed. The LVCs were organised according to the degree of frequency in the corpora under the label frequency: high, medium, and low. Also, all examples were classified according to the following grammatical factors:

(i) determination (yes/no),

(ii) modification (different classes),

---

[24] The database is made available on the OSF Platform for the scientific community. Open access at: https://osf.io/nbje6/?view_only=57069d42f0dc48deb3c6bb25840fbe4e

[25] LVCs as those in (3) with a preposition introducing the nominal −when they generally do not take a preposition, unlike *tomar en consideración* 'take into consideration'− have not been studied yet. However, the role that this preposition plays in the meaning of the LVC requires further syntactic-semantic investigation. From the corpus instances found in CORPES XXI, it seems that these prepositions are included to express some kind of repetition in certain areas of Latin American Spanish.

(iii) number of the noun (singular/plural).

Some methodological limitations must be pointed out. A certain margin of error is still possible since after the automatic retrieval all irrelevant instances must be manually removed, and this selection may be subject to bias or assumptions on the part of the researcher. 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 found in my database are compensated with a high number of instances, and the procedure of decision was informed by previous literature. All remaining errors are the researchers' own responsibility.

In the case of the first grammatical variable, *determination* is understood in a broad sense: the delimitation of the extension of the noun phrase. Traditionally, articles (4), demonstratives (5), possessives (6), numerals (7) and quantifiers (8) are classified as determiners (RAE/ASALE, 2009: 1.9r; GIEC, 2016: 11.1; Huddleston & Pullum, 2002; Eisenberg et al., 2005); and they have all been analysed as determiners in my database for the four languages.

(4) a. The big teeth take *the big bites*.   (COCA, 2006)

b. dann sollte er *einen Hinweis* geben   (DWDS, 2005)

'then he should give an indication'

c. un pare no se sent mai avergonyit per haver *pres la decisió* d'abandonar el seu fill.   (CTILC, 2011)

'a father never feels ashamed of having made the decision of abandoning his child'

d. No había tomado *la decisión*,   (CORPES XXI, 2001)

'He hadn't made the decision'

(5) a. We're not going to *make this mistake* again;   (COCA, 2012)

b. die *diese Fehler machen*.   (DWDS, 2013)

'that make this mistake'

c. Cal anar amb compte a l'hora de *fer aquesta anàlisi*   (CTILC, 2015)

'One must be cautious when making this analysis'

d. es importante *hacerse esta pregunta*   (CORPES XXI, 2002)

'it is important to ask this question'

(6) a. That's President Obama today, preparing to *give his farewell speech* this upcoming week in Chicago (COCA, 2017)

b. wir haben uns natürlich *unsere Gedanken gemacht*, (DWDS, 2006)

'sure, we have made our thoughts'

c. als ciutadans de la vila que ens van *donar el seu suport* a les urnes.

(CTILC, 2006)

'to the village citizens who have us their support in the ballots'

d. — ¿Da usted *su permiso*? (CORPES XXI, 2002)

'Do you give your permission?'

(7) a. they will give you *two different answers* (COCA, 2012)

b. Dabei machen sie *zwei Fehler* auf einmal: (DWDS, 2018)

'They make two mistakes at once:'

c. feia *milers de dibuixos* diferents dins de les finestretes fosques de les tauletes dels esperits. (CTILC, 2017)

's/he made thousands of different drawings inside the dark windows of the spirits' tablets.'

d. debe dar *tres clases* (CORPES XXI, 2006)

'he must give three lessons'

(8) a. Paige and I will come to your wedding, but make *no mistake* (COCA, 2016)

b. Ich kann darauf *keine Rücksicht nehmen*. (DWDS, 2004)

'I cannot take it into (any) consideration'

c. però no semblava pas que hi fessin *cap referència*. (CTILC, 2003)

'but it didn't seem that they would make any reference to it'

d. en Colombia no dan *ninguna ayuda* para la cultura. (CORPES XXI, 2003)

'in Colombia no help is given to culture'

In addition, other elements which are not prototypically determiners have been classified within this label because of their function within the noun phrase: quantifiers such as *more, mehr, més, más* (9), as they also modify the nominal (RAE/ASALE, 2009: §1.9r) although they do not affect their reference or identification but rather their quantification (GIEC, 2006: §12.2); some functional adjectives such as *same, gleich, mateix, mismo* (10), which have previously been classified as indefinite determiners in Romance languages (Charnavel, 2010; Bosque, 2012; Lopes, 2018) as well as identity adjectives in English (Oxford, 2010).

(9) a. You should *take more naps* on those sofas in the Oval.   (COCA, 2014)

b. Wir müssen auf die Natur *mehr Rücksicht nehmen.*   (DWDS, 2002)

'We should take nature more into consideration'

c. i no li va caldre *donar més explicacions.*   (CTILC, 2008)

'and he did not have to give more explanations'

d. Martínez recomendó hacer *más ejercicios.*   (CORPES XXI, 2001)

'Martínez recommended doing more exercises'

(10)   a. you may meet with many books which will *give you the same advice*

(COCA, 2012)

b. ich habe ihnen allen *die gleiche Antwort gegeben,* mit einem einzigen Wort: Stimmt!   (DWDS, 2003)

'I gave you all the same answer with a single word: correct!'

c. Però això no vol dir que discriminem ningú i que cap treballador cobri menys que un altre per *fer la mateixa feina.*   (CTILC, 2005)

'But that doesn't mean that we discriminate against anybody nor that any workers earns less than another to do the same work.'

d. Más temprano que tarde el resto de la clientela terminó tomando *la misma determinación.*   (CORPES XXI, 2003)

'Sooner than later the rest of the costumers ended up taking the same determination.'

Other complex determiners have also been included in the determination label in my dataset, such as *kind of/sort of* in English (11) following Brems (2011), among others, who analyse these as complex determiners.

(11)   a. but it will take *that kind of action*.   (COCA, 2019)

b. But the President appointed me to *make these sort of calls*.

(COCA, 2010)

With regard to the variable *modification*, several classes of modification have been included. As previously described in Chapter 2, Levin & Ström Herold (2015) examine the modification of LVCs in Germanic languages (English, German and Swedish), and they find that adjectival modification is the most frequent modification of the nominals in LVCs. In contrast, other types of modification such as clausal modifiers represent a residual percentage of the instances in their study, between 1% and 5% in English instances and 0% and 2% in German (Levin & Ström Herold, 2015: 24).

However, due to the different tendencies of nominal modification between Germanic and Romance languages, where Romance languages have a greater prevalence for prepositional phrases (Bonet & Solà, 1986; Cremades, 2016), the present study has established the following labels for the classification of variable modification: adjectival modification for both prenominal and postnominal adjectives (12), noun classifiers (13), prepositional phrases (14), genitive modification (15) and relative clauses (16).

(12)   a. and *take big, long walks* with the dogs.   (COCA, 2019)

    b. daß sie *einen schrecklichen Fehler gemacht* haben könnte.

        (DWDS, 2002)

    'that she could have made a terrible mistake'

    c. però sense aconseguir *donar-li cap explicació satisfactòria*.

        (CTILC, 2017)

    'but without succeeding in giving him any satisfactory explanation'

    d. el jefe de Estado les dio "*instrucciones muy precisas*"

        (CORPES XXI, 2001)

    'the head of State gave them very precise instructions'

(13)   a. You fail to *make your mortgage payment* by the due date.

        (COCA, 2012)

    b. Ich wollte ihnen nur *einen gute Nacht Kuss* geben.

        (DWDS, 2006)

    'I only wanted to give you a good night kiss'

(14)   a. Atheism itself make no *mention of feminism or misogyny*.

        (COCA, 2012)

    b. Sie machte es, als sie *Angaben über Wallowski* machte.

        (DWDS, 2010)

    'She made it, when she gave details about Wallowski'

    c. fas *la feina del pis* i per menjar i per la cambra no m'has de donar res;

        (CTILC, 2005)

    'you do the housework and the cooking, and you don't need to pay for the room'

    d. se tomó *un descanso de seis meses*.   (CORPES XXI, 2004)

    'he took a break of six months'

(15)   a. We'll *take a day's hike* on the Appalachian Trail.   (COCA, 2016)

b. Karl Schlögel hat uns *eine Beschreibung der Mitte* gegeben.

(DWDS, 2001)

'Karl Schlögel gave us a description of Mitte'

(16) a. She'll give her kids *an answer they don't like.*   (COCA, 2011)

b. wenn er Erfahrungen gemacht hat, *die der Autor auch gemacht.*

(DWDS, 2000)

'when he had the experience that the author also had'

c. i els «veterans» em van donar *una pallissa que em va deixar baldat*.

(CTILC, 2011)

'and the seniors gave me a beating that left me smashed'

d. quienes hicieron *el análisis que determinó esa decisión*,

(CORPES XXI, 2001)

'those who made the analysis which determined that decision'

It is important to note that the countability of the head noun is relevant when approaching its determination and modification properties. For Husić (2020), abstract nouns (such as eventive nouns) are usually left out of the count/mass distinction because cumulativity, divisiveness and atomicity are not found when the reference is abstract. Therefore, she proposes that countability in abstract nouns is related to their semantic category. Specifically, telic eventualities are regarded as count (whether they are inherently telic, i.e. *death*, or the result of a modification of a process, i.e. *transplant*), states are mass (and they resist mass to count coercion, i.e. *love*), and processes are flexible regarding countability (i.e. *inquiry*), in the sense that the core meaning is atelic and unbounded, but they can shift to telic eventualities. A similar proposal for Spanish non-deverbal nominalizations is proposed by Resnik (2010: 212), who distinguishes between mass eventive nouns (i.e. *pánico* 'panic', *rabia* 'anger', *caos* 'chaos', *frío* 'coldness', etc.), and ambiguous eventive nouns between countable (*cada crisis* 'every crisis', *cualquier fiesta* 'any party') and uncountable uses (*hay fiesta* 'there is party', *hay guerra* 'there is war'). Determination is key for the mass/noun contrast (Alexiadou, 2001; Kornfeld, 2004; Resnik, 2010), as it has been shown that the absence or presence of a determiner is more relevant than the inherent semantic properties of the nominal, which have no influence beyond the limits of nP (Resnik, 2010: 278).

Taking this into account, the count/mass classification of the nominals selected for this corpus-based study is relevant for the analysis. However, since the focus of this

dissertation are eventive LVCs, the presence of mass abstract nouns is rather limited. Besides the two inherently mass nouns in English (*advice* and *progress*), the rest of nominals are dependent on the absence/presence of the determiner for the mass/noun distinction[26]. Following Rothstein (2010: 346-347), for English and German, the possibility to appear with quantifiers like Eng. *much*/ Ger. *viel* (17) as well as with plural morphology or other determiners (18) proves the duality of such nominals, which can behave both as mass (17) and count (18).

(17) a. It may be helpful to spend some time doing things together which do not require *much speech*. (Google)

  b. die im Urlaub *viel Sport* gemacht haben. (DWDS, 2013)

(18) a. presidents do not necessarily lose their ability to give *a good speech*. (COCA, 2016)

  b. weil sie *keinen Sport* machen kann? (DWDS, 2013)

Following Resnik (2010), for Romance languages, the possibility to appear with quantifiers like Cat. *molt de*/ Sp. *mucho* (18) as well as with other determiners or plural morphology (19) proves that they can behave both as mass (19) and count (20).

(19) a. El govern francès sempre ha donat *molt de suport* al comerç de la roba. (CTILC, 2009)

  'The French government has always given a lot of support to the clothing trade'

---

[26] While the majority of nominals are count (i); there is also a group of nominals belonging to the intermediate group which can be count or mass depending on the presence of the determiner (ii).

  (i) a. answer, try, hug, smile, kiss, nod, push, sigh, example, step, walk, shower, bite, nap, bath, hike, stroll, swim, picture, decision, mistake, call, claim, contribution, payment, assumption, mention. (English)
  b. Hinweis, Antwort, Unterricht, Erlaubnis, Kuss, Warnung, Beschreibung, Versprechen, Rat, Abschied, Einfluss, Abstand, Rücksicht, Einsicht, Anteil, Anstoß, Gedanke, Sorge, Angabe, Anfag, Ausführung, Beobachtung, Foto, Fehler. (German)
  c. resposta, volta, explicació, empenta, definició, ajut/da, bes/ada, cop, conferència, bufetada, pallissa, decisió, consciència, iniciativa, molèstia, represàlia, nota, pregunta, visita, viatge, petó, dibuix, abraçada, aclariment, fotografia. (Catalan)
  d. ayuda, autorización, instrucción, giro, salto, cambio, beso, golpe, vuelta, respuesta, bofetada, decisión, conciencia, iniciativa, descanso, baño, acuerdo, represalia, atajo, nota, fotografía, pregunta, viaje, declaración. (Spanish)
  (ii) a. speech, notice, chase, breath, risk, drink, advantage, action, choice, adjustment, announcement, accusation, profit. (English)
  b. Platz, Eindruck, Erfahrung, Hoffnung, Sport. (German)
  c. suport, consell, permís, possessió, precaució, bany, impuls, determinació, partit, distància, referència, esforç, ús, anàlisi, feina. (Catalan)
  d. paso, consejo, permiso, clase, precaución, determinación, distancia, esfuerzo, uso, análisis, énfasis, crítica, colección, carrera, broma, ejercicio. (Spanish)

b. A los de la pandilla yo los regañaba así, a boca, y les daba *mucho consejo.* (CORPES XXI, 2002)

'I used to scold the gang like that, with my mouth, and I gave them a lot of advice.

(20)   a. i *donà un suport decidit* a la Generalitat republicana. (CTILC, 2015)

'and gave firm support to the Republican Generalitat.'

b. Un padre que *da consejos* más que padre es un enemigo.

(CORPES XXI, 2003)

'A father who gives advice is more of an enemy than a father.'

After the manual classification, the final database is described in Table 14: 7,243 instances of English LVCs; 3,476 examples of German LVCs; 2,985 instances of Catalan LVCs, and 7,630 instances of Spanish LVCs. The total data sample that is taken as the empirical basis of this study comprises 21,334 instances. The divergences in the total amount of examples analysed for each language are consistent with the differences discussed earlier regarding the size of the corpora (seen in Table 5).

	English	German	Catalan	Spanish	TOTAL LV
Give | Geben | Donar | Dar	2,026	995	931	2,842	6,794
Take | Nehmen | Prendre |Tomar	2,530	1,120	513	2,358	6,521
Make | Machen | Fer | Hacer	2,687	1,361	1,541	2,430	8,019
TOTAL LANG	7,243	3,476	2,985	7,630	21,334

Table 14. Distribution of the number of LVCs of the database for each language

More specifically, the final database of *give*-LVCs instances consists of 6,794 tokens of LVC with some differences in the distribution among the four languages: Spanish and English provide the highest number of examples above 2,000 each, and German and Catalan only above 900. For *take*-LVCs, the final database consists of 6,521 instances with a distribution of more than 2,500 for English and 2,000 for Spanish, more than 1,000 for German and only above 500 for Catalan. Finally, *make*-LVCs provide the highest number of tokens with 8,019. In this case, English presents more than 2,500 instances and Spanish slightly below 2,500, while both German and Catalan have a total of more than 1,000.

3.3. Data analysis

This section reports the data analysis of the corpus-based study which constitutes the empirical basis for this dissertation. The main quantitative analysis of the results will focus on descriptive statistics which mainly describe the data through counts and percentages. When comparing groups, languages and results, some inferential statistics will be introduced which allow to make inferences about the generalization of observed patterns, such as the chi-square test (also used in previous cross-linguistic studies about LVCs, Levin & Ström Herold, 2015). Some basic calculations have been carried out via MS Excel (through the Analyze Data toolbox), and the chi-square tests have been calculated on the Lancaster Stats Toolbox online (Brezina, 2018)[27]. Such significance tests allow us to determine whether a finding is the result of a difference between two (or more) items, or whether it is just due to chance. Thus, when the difference is significant ($p < 0.001$), it shows that there is a high degree of certainty that the difference is a reflection of variation.

Egbert et al. (2020) argue in favour of descriptive statistics being sufficient statistical methods when used to appropriately describe the data, while inferential statistics enhance the interpretability of the statistical results. At this point it is important to take into account that the null hypothesis significance testing (i.e. chi-square test) is sensitive to sample size, that is, the larger the sample, the more likely it will be to find a statistically significant result. In order to control for this, other statistical measures such as the effect size (whether it is small, medium or large) will be provided (following Brezina, 2018).

After all, following Egbert et al. (2020, 2022), the output of the statistical tests will not be the main contribution of the study, and they will be complemented and enriched through linguistic analysis and description of the actual data. For this reason, the quantitative results will be put forward together with descriptive and qualitative analysis of the data in the upcoming subsections: *give*-LVCs (§3.4), *take*-LVCs (§3.5) and *make*-LVCs (§3.6). For each verb, the following subsections have been included: determination and modification of the NVE. More specifically, the general results of the modification are also specified for the variable degree of frequency, followed by the results of each particular LVC in English, German, Catalan and Spanish.

---

[27] Available at: http://corpora.lancs.ac.uk/lancsbox

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study

In order to confirm the empirical trends attested, a binary regression analysis will be calculated for each language with the aim to observe the relation between the modification, and the rest of the variables under study. With this methodological analysis, relevant predicting factors will be identified and discussed in §3.7.

Finally, the conclusion and final remarks are provided in §3.8.

3.4. Quantitative results: the LV *give*

As mentioned in the preceding section, the LV *give* is quite stable in terms of frequency in English (Butt, 2010; Elenbaas, 2013), which contrasts with the German verb *geben,* which is not among the most frequent LVs (Bruker, 2011). In the Romance languages, Spanish *dar* and Catalan *donar* are among the most common verbs, with the difference that this LV is the most frequent verb in Spanish (Sánchez Rufat, 2015), while it is not considered the most basic LV in Catalan, because this language favours *fer* ('make') (as pointed out in Rosselló, 2002; Ginebra & Navarro, 2015; GIEC, 2016: 17.2.2).

The sample was obtained by searching for the collocation of the LV *give* (and its corresponding translation in each language: Ger. *geben,* Cat. *donar* and Sp. *dar*) and the specific nominal element with which it builds an LVC. The final selection of *give*-LVCs to be analysed for each language was set to a maximum of 15 LVCs per language, with a range from highly frequent LVCs to less frequent (see Table 15, repeated from above for convenience).

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical 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')

Table 15. LVCs analysed in the four languages under study

The final database of *give*-LVCs instances consists of 6,789 tokens of LVC with some differences in the distribution among the four languages: Spanish provides 2,842 tokens, English presents a slightly lower number with 2,030, while German and Catalan are below 1,000 instances (995 for German, and 931 for Catalan) (see Table 16).

	English	German	Catalan	Spanish	TOTAL LV
Give | Geben | Donar | Dar	2,030	995	931	2,842	6,798

Table 16. Distribution of the number of LVCs of the database for each language

As seen in previous chapters, both determination and modification in LVCs are flexible. Corpus data allow to establish certain patterns as well as comparisons between and within language families. The following subsections focus on determination (§3.4.1) and modification (§3.4.2) in *give*-LVCs. Results will be presented for each language and each particular LVC through corpus linguistics methods, such as frequency and chi-square tests to compare observed results with expected results. After this, results will be analysed and compared with previous corpus-based studies in the discussion in §3.4.3.

3.4.1. Determination in *give*-LVCs

Figure 1 shows that determination (DET) is predominant in *give*-LVCs in the Germanic languages under study, English (81.74%) and German (77.99%), and to a lesser extent in the Romance languages, Spanish (63.27%) and Catalan (51.02%). When the focus is placed on bare nominals, the Germanic languages show opposite tendencies: in English, bare singulars (BS) appear more often (11.06%) than bare plurals (BPl) (7.20%); while bare singulars in German are less frequent (8.44%) than bare plurals (13.57%). A similar trend is found in the Romance languages: in Catalan, bare singulars are more common (31.47%) than bare plurals (17.51%), while bare singulars in Spanish appear less (15.13%) than bare plurals (21.60%) in my sample. This distribution is statistically significant when all languages are compared ($\chi 2(6)= 493.55$, $p<.0001$, Cramer's V = 0.191), as well as within language families: when Germanic languages are compared statistical tests prove a significant difference ($\chi 2(2)=34.67$, $p<.0001$, Cramer's V = 0.107, small effect), and also do Romance languages ($\chi 2(2)=120.89$, $p<.0001$, Cramer's V = 0.179, small effect).
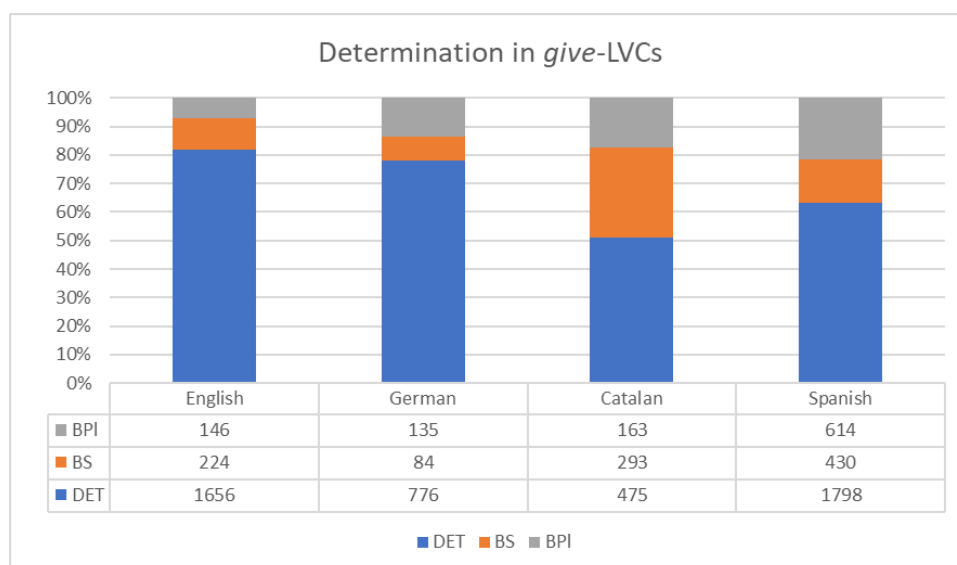
Figure 1. Distribution of the determination in give-*LVCs*

Although these differences could be traced back to the choice of LVCs analysed for each language and their different combinatorial possibilities with bare nominals, the distribution of obligatory and optional determiners in *give*-LVCs (Table 17) is only explanatory in the case of English and Spanish.

	Optional DET	Obligatory DET
give	Advice, example, notice, chase	Speech, answer, try, hug, smile, kiss, nod, push, sigh
geben	Antwort ('answer'), Erlaubnis ('permision'), Rat ('advice'), Beschreibung ('description'), Unterricht ('lesson'), Versprechen ('promise'), Warnung ('warning')	Hinweis ('hint'), Kuss ('kiss')
donar	Suport ('support'), resposta ('answer'), consell ('advice'), permís ('permision'), empenta ('push'), conferència ('conference'), ajuda ('help'), explicació ('explanation')	Cop ('blow'), volta ('stroll'), definició ('definition'), bufetada ('slap'), pallissa ('beating'), bes/ada ('kiss')

| dar | Autorización ('authorization'), ayuda ('help'), vuelta ('stroll'), paso ('step'), respuesta ('answer'), clase ('lesson'), cambio ('change'), consejo ('advice'), instrucción ('instruction'), permiso ('permision') | Beso ('kiss'), golpe ('blow'), salto ('jump'), giro ('turn'), bofetada ('slap') |

Table 17. Distribution of obligation and optionality of a determiner in *give*-LVCs

In English, the majority of LVCs require a determiner, while in Spanish most LVCs have an optional determiner. Therefore, differences in determination percentages can be explained in relation to the possibility of accepting bare nominals in LVCs in each language. In contrast, the distribution in Table 17 is not as explanatory for German and Catalan. While determiners are minimally obligatory in German, and almost all LVCs have an optional determiner, the presence of determination is very high in the overall results (see Figure 1). In contrast, Catalan has an even distribution of LVCs with obligatory determiner (n=7) and with an optional determiner (n=7), but the presence of the determiner in the overall results in my sample are much lower than those of German.

The distribution in Table 17 is aligned with the different tendencies of both languages in the nominal domain, with Germanic languages prioritizing DPs in the object position and Romance languages being more flexible in accepting bare nominals in this position (Chierchia, 1998; Longobardi, 2001).

Figure 2 shows that English *give*-LVCs follow the tendency presented in Figure 1 with a very general prevalence of determination. A more detailed analysis of each lemma shows that only *give notice* and *give chase* present a clearly divergent proportion: *give notice* has a majority of instances combined with the bare singular (62.87%), while the case of *give chase* is even higher with 97.80% of bare singulars (and only 2 instances of a determined nominal). It is important to note here that, besides these two LVCs, only *give advice* shows a noticeable presence of bare singulars within its examples (16%), because it is the only inherently mass noun, whereas the rest of LVCs do not combine with bare singulars in the NVE position in any case, or only in 1 instance in *give example*. Therefore, three LVCs (*give chase, give notice* and *give advice*) influence the general

results in Figure 1 which showed a preference for English *give*-LVCs to select bare singulars, contrasting with the tendency of German to select bare plurals.
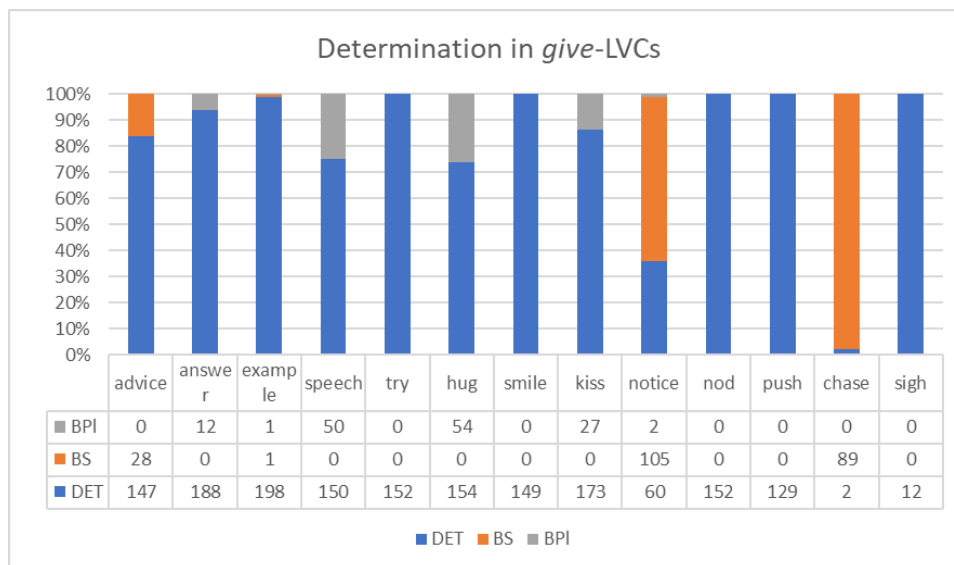
Figure 2. Distribution of the determination in English give-*LVCs per lemma*

The fact that *advice* is a mass noun influences the absence of bare plurals (n=0). However, *notice* and *chase* are both count nouns, but they behave differently than the rest of nouns in *give*-LVCs, as they show a clear preference for the NVE to be a bare singular (in 195 instances and 89, respectively), as in (21).

(21)   a. he did not *give* either candidate early *notice* of his endorsement.

(COCA, 2012)

b. I presume you declined, since he left and you were driven to give chase.

(COCA, 2018)

Figure 3 shows that determination is prevalent in German *geben*-LVCs in all cases but one: *Unterricht geben* has more instances of bare singulars (63.27%) than determined NVEs (36.73%), as in (22). Likewise, *Hinweis* and *Kuss* combine only with bare plurals. The analysis of the different lemmas selected for the analysis also corroborates the tendency observed in Figure 1 of German *geben*-LVCs to prefer bare plurals than bare singulars, as shown in all LVCs with the only exceptions of *Unterricht geben* and *Rat geben*.
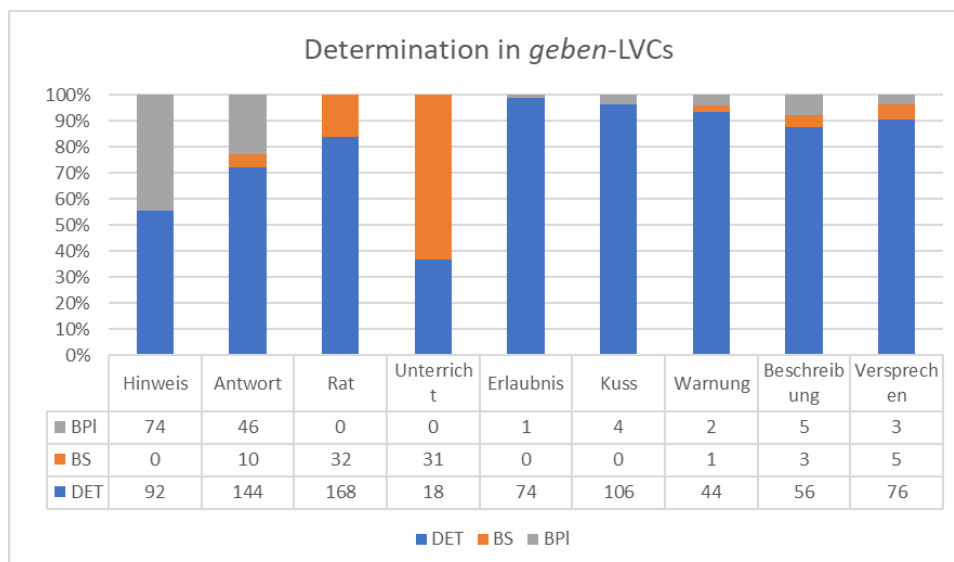
Figure 3. *Distribution of the determination in German* geben-*LVCs per lemma*

(22)   a. Jeden Tag drei Stunden lang soll er mir *Unterricht geben*.

                                                                    (DWDS, 2006)

'He should give me classes every day for three hours'

b. dass jene Lehrer, die streiken und *keinen Unterricht geben*, bestraft werden.   (DWDS, 2000)

'that those teachers who strike and do not give any lesson will be punished.'

The analysis of the determination in Catalan *donar*-LVCs (Figure 4) shows a clear tendency to appear with nominals introduced by a determiner with three exceptions: *donar suport* (7% of determined NVE vs. 93% of bare singulars), *donar permís* (26.47% of determined NVE vs. 70.59% of bare singulars and 2.94% of bare plurals) and *donar ajut/da* (36.84% of determined NVE vs. 47.38% of bare singulars and 15.78% of bare plurals).
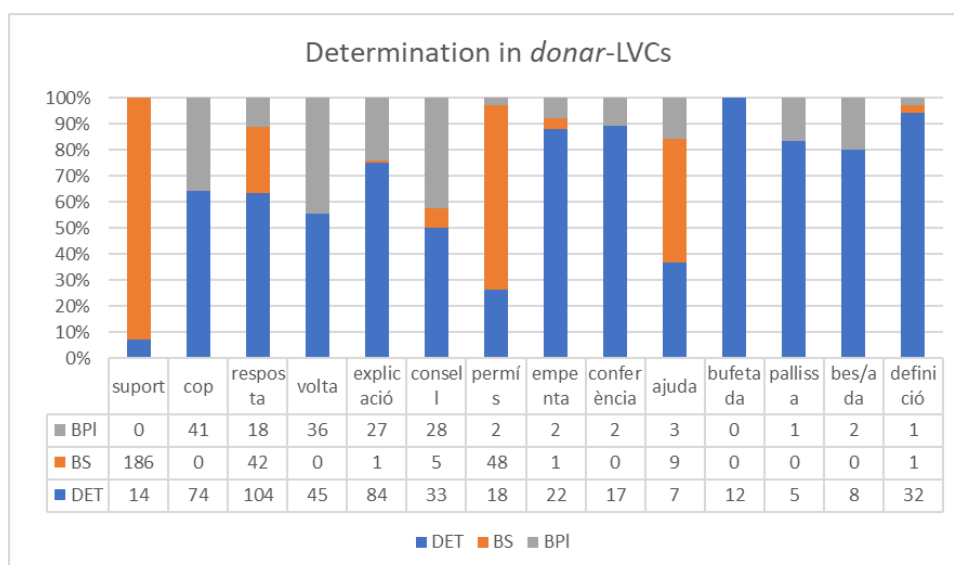
Figure 4. Distribution of the determination in Catalan donar-*LVCs per lemma*

For these three LVCs, the trend is the opposite. They select bare singulars over bare plurals, as illustrated in (23). It is also important to note that all LVCs with an obligatory determiner (established in Table 17) combine only with determined nominals and bare plurals, with the only exception of *donar definició* that appears with a bare singular. However, its being part of a coordination with another bare singular may have influenced the absence of the determiner (24).

(23)   a. entre els partits que *donen suport* a l'executiu català.

(CTILC, 2013)

'among the parties which give support to the Catalan government'

b. però ella s'ha instal·lat en un racó al costat de la llar i assegura que no es mourà d'allà fins que no li *donin permís* per entrar aquí.

(CTILC, 2001)

'but she has settled in a spot close to the fireplace and claims that she will not move from there until she is given permission to come in'

c. que *donaven ajuda* en la malaltia i la mort i que no eren simples confraries de caràcter religiós.   (CTILC, 2003)

'which gave help in sickness and death and were not mere religious associations'

(24)   a. De seguida que *dono nom i definició* als meus sentiments,

(CTILC, 2004)

'As soon as I give name and definition to my feelings'

86

An examination of Spanish *dar*-LVCs in my sample shows that almost all of them have nominal elements introduced by a determiner, but there are some exceptions, as displayed in Figure 5: *dar vuelta* with only a 41% of instances, *dar clase* with 17%, *dar instrucción* with 82%, and *dar permiso* with 28.50%.

Determination in *dar*-LVCs

	vuelta	paso	respuesta	beso	golpe	clase	salto	cambio	consejo	instrucción	giro	permiso	ayuda	bofetada	autorización
BPl	69	17	27	12	38	131	53	4	78	130	18	6	19	12	0
BS	49	31	67	0	0	35	0	0	4	8	0	137	63	0	36
DET	82	152	106	188	161	34	147	57	118	62	164	57	118	188	164
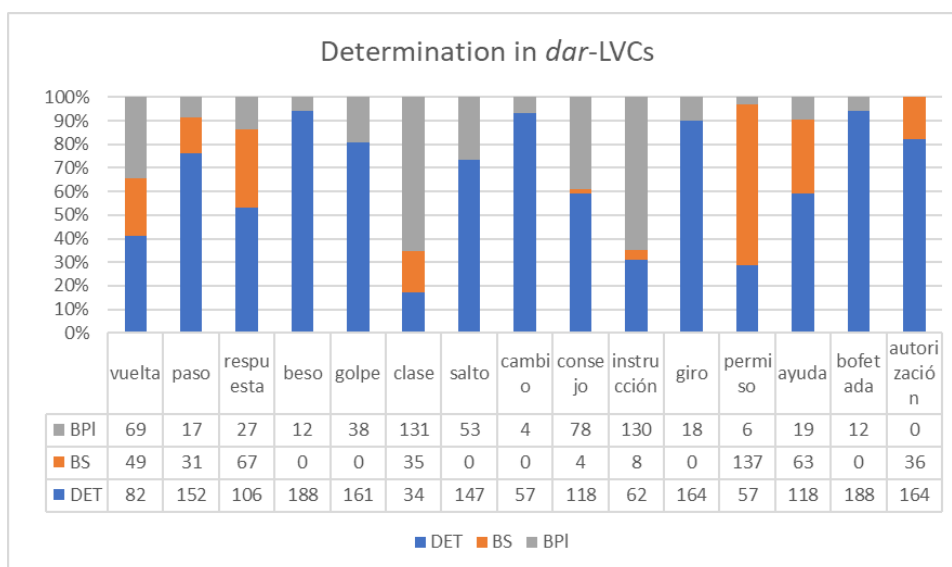
Figure 5. Distribution of the determination in Spanish dar-*LVCs per lemma*

The kind of predominant bare nominals in each of these lemmas is not the same in every case: most select bare plurals more often, such as *dar vuelta* (34.50% bare plurals vs. 24.50% bare singulars), *dar clase* (65.50% bare plurals vs. 17.50% of bare singulars), and *dar instrucción* (65% bare plurals vs. 4% bare singulars),while *dar permiso* combines most often with bare singulars (68.50% bare singulars vs. 9.50% of bare plurals), as illustrated in (25).

(25) a. un montón de finas lonjas de carne atravesadas por un fierro *dan vueltas* alrededor del fuego.   (CORPES XXI, 2001)

'a lot of thin slices of meat pierced by an iron turn around the fire'

b. Hoy mismo he recibido una propuesta para *dar clases* en el Colegio Ruso,   (CORPES XXI, 2007)

'Today I have received a proposal to teach in the Russian School'

c. Cuando ha terminado de escribir, le *da instrucciones* sobre cómo y a quién pedir la consulta con el cirujano a la entrada del consultorio.

(CORPES XXI, 2001)

'When he has finished writing, he gives him instructions about how and who to ask an appointment with the surgeon in the doctor's office lobby.

d. La banda de 'punk-rock' vende sus CD a tres euros y *ha dado permiso* para piratearlos en Europa y Latinoamérica. (CORPES XXI, 2002)

'The punk-rock band sells their CD for three euro and has given permission to pirate them in Europe and Latin America'

Also, LVCs with an obligatory determiner (classified in Table 17) only combine with DPs and bare plurals, which are more flexible in an object position than bare singulars, as *dar beso, dar golpe, dar salto, dar giro,* and *dar bofetada* in Figure 5, and examples in (26).

(26)   a. Agustín la encontró cuando le *daba besos* en todo el pie.

(CORPES XXI, 2001)

'Agustín found her when he was giving her kisses on her feet'

b. había hablado con científicos que "se *dan golpes* contra la pared" ante esta cuestión.   (CORPES XXI, 2012)

'He had talked with scientists who "beat the wall" regarding this matter'

c. Siempre subía y bajaba esta escalera *dando saltos*,

(CORPES XXI, 2001)

'He always jumped up and down the stairs'

d. Josef le *da giros* a la cabeza.   (CORPES XXI, 2002)

'Josef turns his head around'

e. pero sin posibilidad de clavar uñas o *dar bofetadas*.

(CORPES XXI, 2007)

'but without any possibility to stick fingernails or slap'

Interestingly, bare nouns are more common in Spanish and Catalan (as seen in Figure 1), although they are less frequently modified in these languages, as Figure 6 shows[28], while Germanic languages present higher proportion of modified bare nominals. In English, 51.34% bare singulars appear with some kind of modification versus 36.99% of bare plurals. The opposite tendency is found in German, with 30.95% of bare singulars versus 49.63% of bare plurals with modification. In both languages, the differences in number for the presence of modification are significant ($\chi2$ (1)=7.34, p=0.006, Cramer's V=0.141, small effect; and $\chi2$ (1)=7.39, p=0.006, Cramer's V = 0.184, small effect,

---

[28] All kinds of modification have been included in this section. A more detailed analysis of subclasses of modification will be presented in §3.4.2.

respectively), although in opposite directions. By contrast, Romance languages have much lower percentages of modification in bare nominals, and they share the fact that bare singulars (9.56% in the case of Catalan data, and 12.79% in Spanish) are less modified than bare plurals (35.58% in Catalan, and 25.90% in Spanish). The differences for these languages are also statistically significant ($\chi 2$ (1)=46.36, p<.0001, Cramer's V=0.319, medium effect; and $\chi 2$ (1)=26.65, p<.0001, Cramer's V=0.16, small effect, respectively).
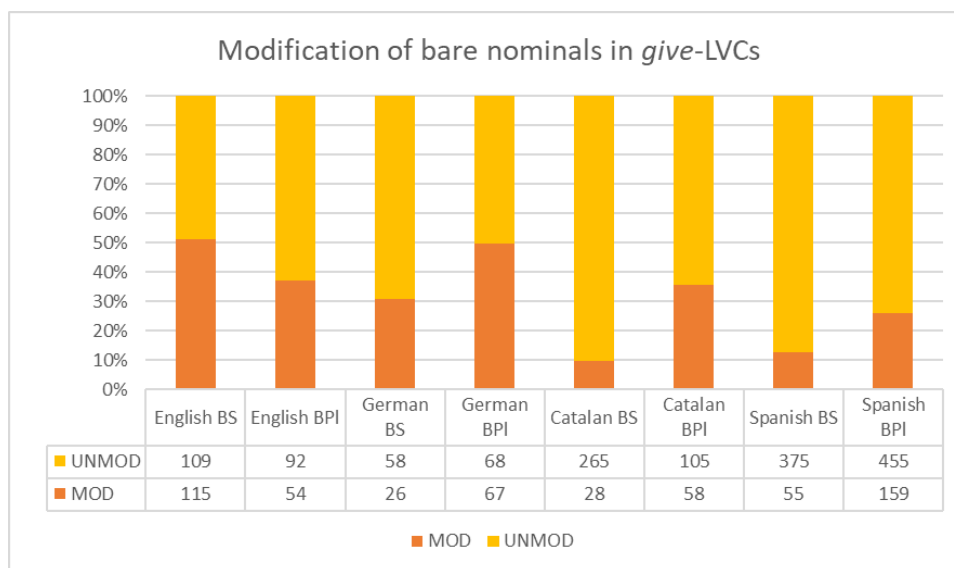
Figure 6. Distribution of the modification of bare nominals in give-*LVCs*

The distribution of determination in *give*-LVCs has been proven to be generalized, since bare nominals are the minority for all languages (Figure 1). However, upon a finer-grained analysis of the sample data, it has been found that *give*-LVCs combine more easily with bare nominals in the Romance languages under study, which simultaneously show a lower tendency to accept modification of these bare nominals, being only more common with bare plurals, as in (27).

(27)   a. Sp. mientras la región comienza a dar *pasos fuertes* en materia de

comercio en línea.   (CORPES XXI, 2001)

'while the region begins to take strong steps in the matter of online

commerce.'

b. Cat. Aquí qui dóna *ajudes puntuals* d'aquest tipus és Càritas.

(CTILC, 2001)

'Who gives episodic help of this kind here is Càritas'

Conversely, the Germanic languages under study have a much lower presence of bare nominals in the object position of *give*-LVCs, but these coappear with modification in more cases (28).

(28)   a. that the staff will give *correct and complete answers*.   (COCA, 2012)

    b. konnten die Lehrer, wie sie selbst erstaunt berichten, nachmittags wieder *vernünftigen Unterricht* geben.   (DWDS, 2006)

    'as she announced amazed herself that the teachers could continue giving proper lessons in the afternoon.'

A more fine-grained analysis of the modification in *give*-LVCs in general is thus needed to draw conclusions that help to account for this correlation in a principled manner. The next section will present an analysis of the different subclasses of modification included in this study. The focus will be mainly on adjectival modification because of its proportional relevance.

3.4.2. Modification in *give*-LVCs

The analysis of the data on modification in *give*-LVCs show 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 7). 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).
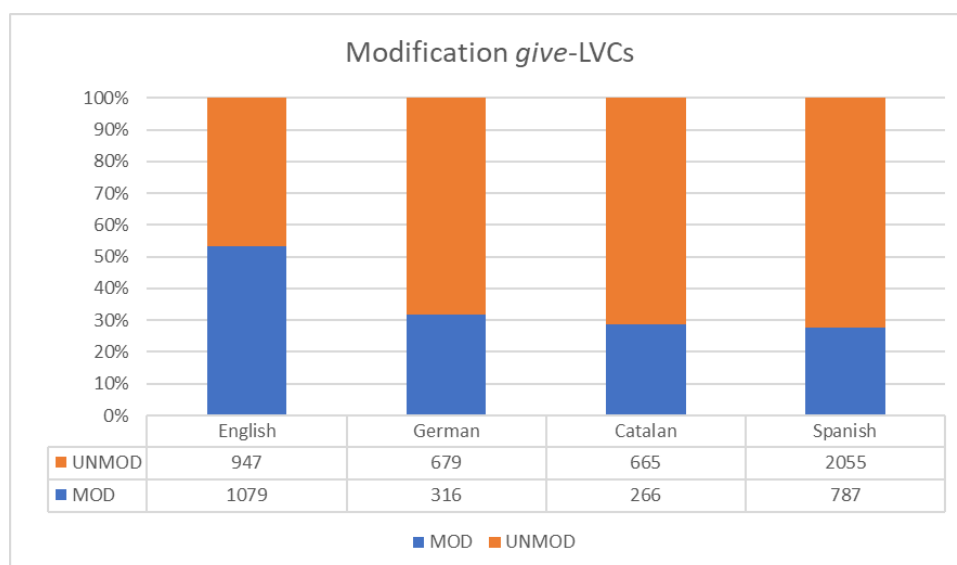
Figure 7. Distribution of modification in give-*LVCs*

The results thus point towards a difference in the tendency to appear modified between English and the other three languages, as shown in Table 18.

Modification	English	German	Catalan	Spanish
Unmod (No)	947 (46.74%)	679 (68.24%)	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

Table 18. Distribution of modification subclasses in *give*-LVCs (all four languages)

When the modification subclasses in Table 18 are taken into account, 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) are only residual in Germanic languages (3.01% in English, 0.30% in German), while they present similar higher percentages in the Romance

91

languages: 6.55% in Catalan *donar*-LVCs (29), and 6.86% in Spanish *dar*-LVCs (30). 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 (as introduced in §3.2.3).

(29) a. David intenta agafar-la, però ella se li gira en contra i comença a *donar-li cops de puny*.   (CTILC, 2001)

    'David tried to catch her, but she turns against him and starts punching him.'

    b. el que tracta de fer és *donar una explicació del context*,

                         (CTILC, 2007)

    'what she tries to do is to give an explanation of the context'

    c. Els buròcrates *donen permís de vida* als cors de la gent de cor buròcrata,

                         (CTILC, 2015)

    'Bureaucrats give life permission to the hearts of people with a bureaucrat heart'

(30) a. Creo que es una reforma muy tímida y el Gobierno tiene que *dar pasos de gigante* porque va con retraso.   (CORPES XXI, 2010)

    'I think that it is a weak reform and the government has to take gigantic steps because they lag behind'

    b. luego resultó clave para la instauración de la República, al *dar instrucciones de neutralidad* a la Guardia Civil. (CORPES XXI, 2001)

    'later it was key for the establishment of the Republic, when giving neutrality instructions to the Guardia Civil'

    c. Carmelo no se sintió autorizado para *dar un giro de ciento ochenta grados* y defender el derecho a la Felicidad (CORPES XXI, 2003)

    'Carmelo did not feel entitled to take a 180 degree turn and defend the right to happiness'

It can be concluded that general differences in the modification of the NVE in *give*-LVCs are statistically validated. Such statistically different trends in modification patterns will be further analysed in the following sections considering the degree of frequency of the LVCs (§3.4.2.1), as well as in relation to each individual LVC for all languages (§3.4.2.2 until §3.4.2.5).

3.4.2.1. Degrees of frequency

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.

Figure 8 shows that English *give*-LVCs with high frequency appear modified in the majority of 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 ($\chi2(2)=40.23$, $p<.0001$, Cramer's V=0.141, small effect).
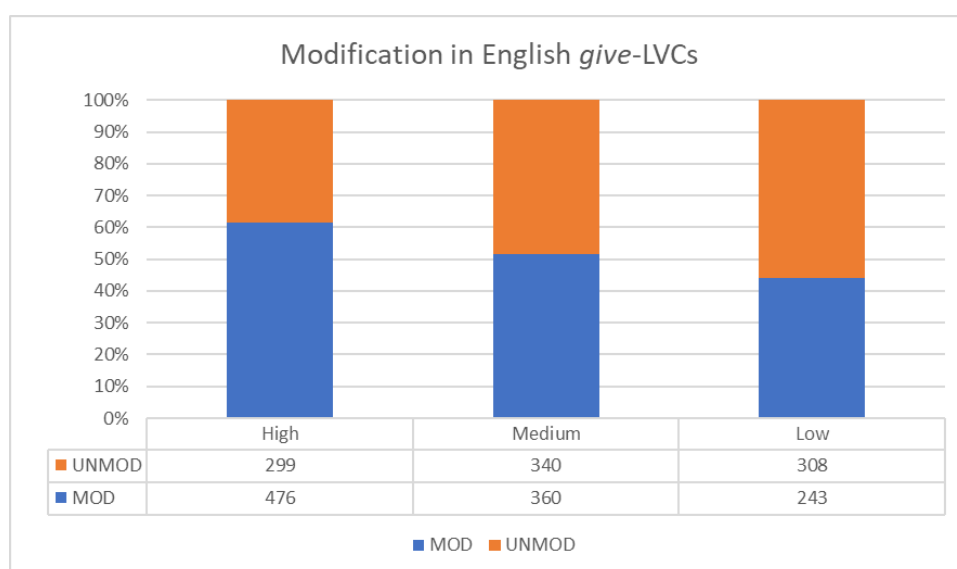
Figure 8. Distribution of modification in English give-*LVCs according to the degree of frequency (high, medium or low)*

Hence, the results for English point to a correlation between degree of frequency and levels of modification: the more frequent an LVC is, the more likelihood that it appears modified. Table 19 shows the distribution of modification subclasses, where the adjectival modification is predominant in the three frequency groups: high (46.19%), medium (46.67%), and in low-frequency LVCs (27.04%).

Modification	High	Medium	Low
Unmod (No)	299 (38.58%)	340 (48.57%)	308 (55.90%)
Adj	358 (46.19%)	327 (46.71%)	149 (27.04%)
PP	42 (5.42%)	-	19 (3.45%)
Rel	19 (2.45%)	7 (1%)	10 (1.81%)
NC	25 (3.23%)	23 (3.29%)	11 (2%)
Gen	-	-	29 (5.26%)
Various	32 (4.13%)	3 (0.43%)	25 (4.54%)
TOTAL	775	700	551

Table 19. Distribution of modification subclasses in *give*-LVCs (according to degree of frequency)

The rest of modification subclasses are residual (3% or below) except for PP in high-frequency LVCs (5.42%), as in (31), and Gen in low-frequency (5.26%), as in (32). The combination of various subclasses (Adj + PP, Adj + Rel, Adj + NC) is also considerable in high-frequency (4.13%) and low-frequency (4.54%).

(31)   a. Since you're so talkative today, why don't you give the class *an example of* Edgar Allen Poe's " Single Effect? " Please.   (COCA, 2017)
b. You *give a speech on the U.S.* getting out of the Persian Gulf and reporters only want to ask   (COCA, 2012)

(32)   Bosses want you to *give two-weeks' notice* before you quit, but they don't do that before they fire you.   (COCA, 2013)

In a similar line, Figure 9 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 degrees of frequency are, again, statistically significant ($\chi2(2)=22.86$, p<.0001, Cramer's V=0.152, small effect).
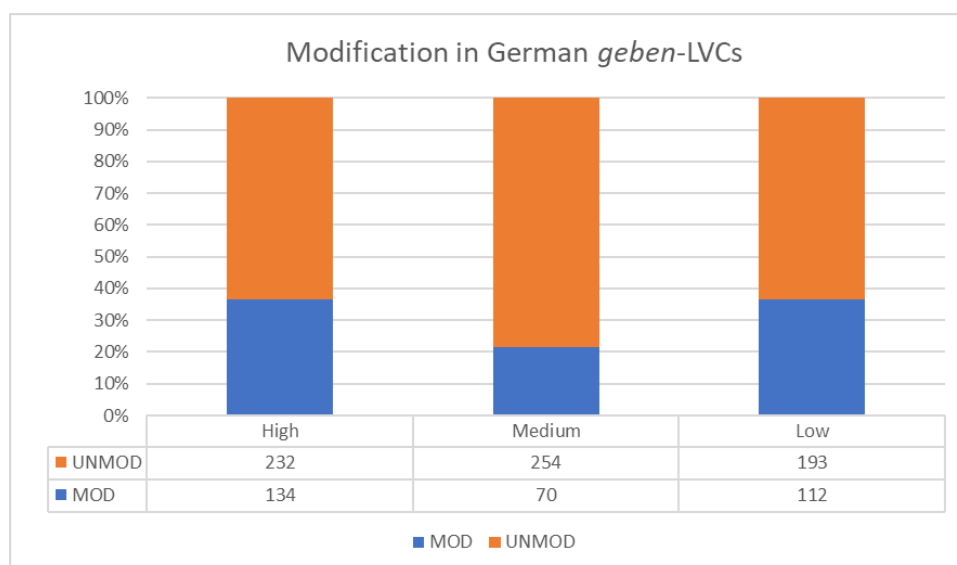
Figure 9. Distribution of modification in German geben-*LVCs according to the degree of frequency (high, medium or low)*

Table 20 shows the distribution of modification subclasses, where adjectival modification is predominant in the three groups: high (34.70%), medium (21.30%) and low frequency (26.23%).

Modification	High	Medium	Low
Unmod (No)	232 (63.39%)	254 (78.40%)	193 (63.28%)
Adj	127 (34.70%)	69 (21.30%)	80 (26.23%)
PP	-	-	3 (0.98%)
Rel	6 (1.64%)	1 (0.30%)	6 (1.97%)
NC	-	-	1 (0.32%)
Gen	-	-	11 (3.61%)
Various	1 (0.27%)	-	11 (3.61%)
TOTAL	366	324	305

Table 20. Distribution of modification subclasses in *geben*-LVCs (according to degree of frequency)

The rest of the modification subclasses are residual in all groups (3% or below), even though there is a relative presence of Gen (3.61%) in low frequency *geben*-LVCs, as well as various modification subclasses (especially, Adj + Gen, in 3.61%), which mostly corresponds to *Beschreibung geben* ('give description')*,* as illustrated in (33).

(33)   a. lzgoev gibt über weite Strecken seiner Analyses *eine Beschreibung der »Pathologie«* des selbstherrschaftlichen Rußland   (DWDS, 2002)

'Izgoev gives a description of the Pathology in the elegant Russian along wide sections of his analysis'

b. Trotz ihres Schrecks war sie so aufmerksam, dass sie der Polizei später *eine genaue Beschreibung* des Jugendlichen geben konnte.

(DWDS, 2011)

'Despite her fright she was very attentive, so later she could give the police an exact description of the young man'

Turning to Romance languages, Figure 10 shows that the distribution of adjectival modification in Catalan *donar*-LVCs presents a different trend that radically contrasts with Germanic languages, which favoured the presence of modification in high-frequency LVCs. 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. 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. 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).
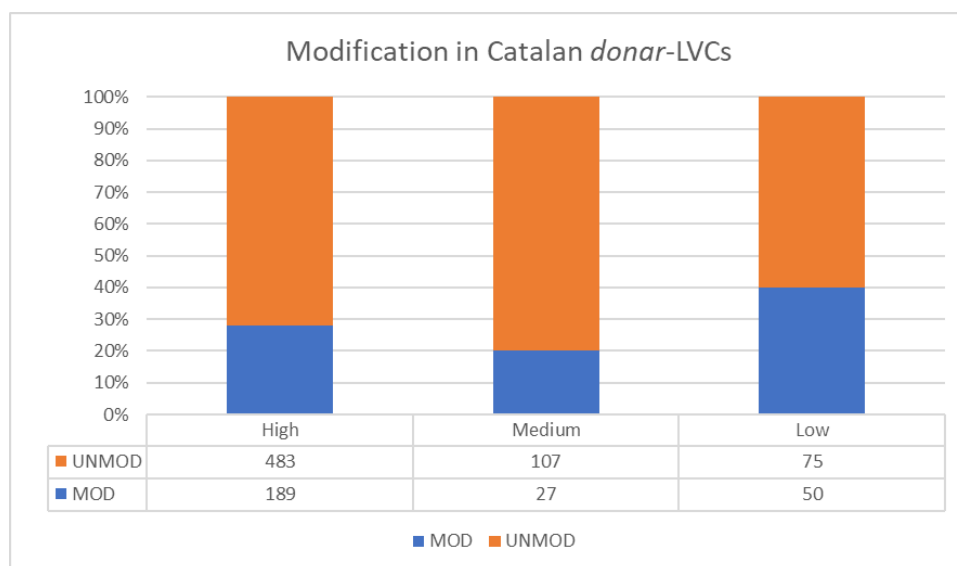
Figure 10. Distribution of modification in Catalan donar*-LVCs according to the degree of frequency (high, medium or low)*

Table 21 shows the distribution of modification subclasses. Once again, adjectival modification is predominant in the three groups: high (18.90%), medium (14.93%), and especially low (26.40%). Regarding the rest of the subclasses, only PP modification is

relevant in all groups (6.70% in high, 5.22% in medium, and 7.20% in low), but also a combination of adjectives and PP in low frequency (4.80%).

Modification	High	Medium	Low
Unmod (No)	483 (71.87%)	107 (79.85%)	75 (60%)
Adj	127 (18.90%)	20 (14.93%)	33 (26.40%)
PP	45 (6.70%)	7 (5.22%)	9 (7.20%)
Rel	10 (1.49%)	-	2 (1.60%)
NC	-	-	-
Gen	-	-	-
Various	7 (1.04%)	-	6 (4.80%)
TOTAL	672	134	125

Table 21. Distribution of modification subclasses in *donar*-LVCs (according to degree of frequency)

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 & McNally, 2011). Bare nominals tend to be modified by relational adjectives (34a), and they only accept qualitative adjectives in restricted contexts (34b).

(34)   a. i *dóna suport tècnic i personal* a la resta d'unitats. (CTILC, 2010)
        'and [it] gives technical and personal support to the rest of the units.'
        b. catalans i balears es donaran *suport mutu*.   (CTILC, 2005)
        'Catalan and Balearic [people] will give mutual support'

These restrictions could have an impact in the results of Catalan *donar*-LVCs which accept the nominal to appear bare. As seen in Table 17, in the sample of Catalan *donar*-LVCs high and medium-frequency LVCs include constructions which accept both determined and bare nominals (35a), as well as other which only accept the nouns with a determiner in singular (35b); while low-frequency LVCs mainly accept determined nominals (36b) and only two instances can combine with bare nominals (36a).

(35)   a. donar (un) suport, donar (una) resposta, donar (un) consell, donar (un) permís
        b. donar *(un) cop, donar *(una) explicació, donar *(una) volta
(36)   a. donar (una) ajuda, donar (una) conferència
        b. Donar *(una) empenta, donar *(una) definició, donar *(una) bufetada,

donar *(una) pallissa, donar *(un) bes/besada

The fact that determined nominals show fewer restrictions in accepting modification might have influenced the results and the differences presented in Figure 10, where the low-frequency *donar*-LVCs show the highest percentages of modification.

For Spanish *dar*-LVCs, Figure 11 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 ($\chi2(2)=45.02$, $p<.0001$, Cramer's V = 0.126, small effect).

	High	Medium	Low
UNMOD	599	549	907
MOD	201	311	275
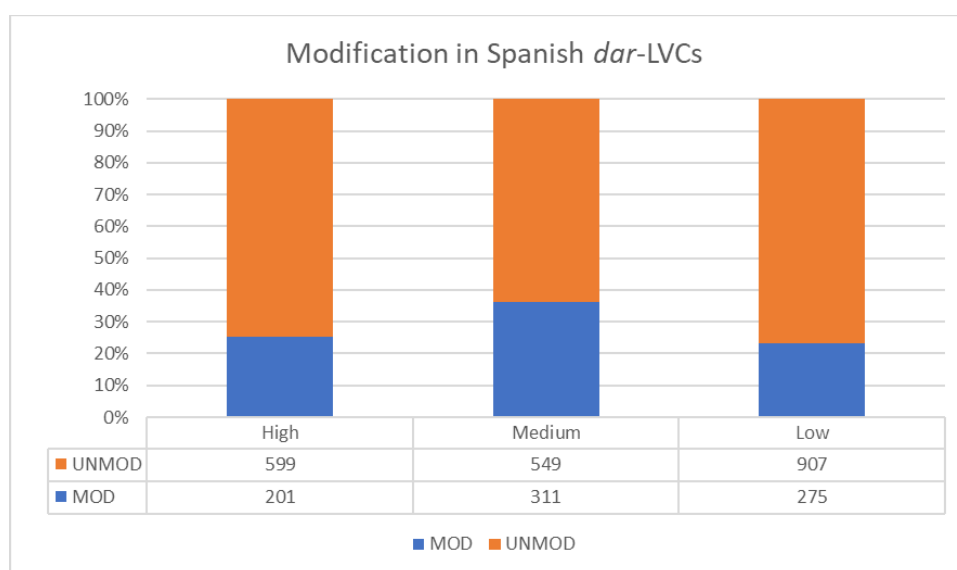
Figure 11. Distribution of modification in Spanish dar-*LVCs according to the degree of frequency (high, medium or low)*

Table 22 shows the distribution of modification subclasses with a predominance of adjectival modification in all three groups in Spanish: high (20%), medium (23.72%), and low (16.24%).

Modification	High	Medium	Low
Unmod (No)	599 (74.88%)	549 (63.84%)	907 (76.73%)
Adj	160 (20%)	204 (23.72%)	192 (16.24%)
PP	30 (3.75%)	96 (11.16%)	69 (5.84%)
Rel	8 (1%)	6 (0.70%)	10 (0.85%)
NC	-	-	-
Gen	-	-	-
Various	3 (0.37%)	5 (0.58%)	4 (0.34%)
TOTAL	800	860	1182

Table 22. Distribution of modification subclasses in *dar*-LVCs (according to degree of frequency)

The rest of the subclasses present residual numbers and proportions (3% or below), except for PP in medium (11.16%) and low-frequency LVCs (5.84%), as illustrated in (37).

(37)   a. porque no dan *clases de educación sexual* en las escuelas.

(CORPES XXI, 2008)

'because they don't teach sexual education in schools'

b. Napster vino a *dar un giro de 180 grados* a todo,

(CORPES XXI, 2001)

'Napster came to take a 180 degree turn to everything'

Regarding the likelihood of *dar*-LVCs to combine with bare nominals, as it is one of the characteristics that sets Romance languages apart from Germanic languages, in my sample there are more *dar*-LVCs which accept both determined and bare nominals in all degrees of frequency (38a) than *dar*-LVCs accepting only determined noun phrases (38b).

(38)   a. Dar (una) vuelta, dar (una) respuesta, dar (un) paso, dar (una) clase, dar (un) cambio, dar (un) consejo, dar (una) instrucción, dar (un) giro, dar (un) permiso, dar (una) ayuda, dar (una) autorización

'take (a) stroll, give (an) answer, take (a) step, give (a) class, make (a) change, give (an) advice, give (an) instruction, take (a) turn, give (a) permission, give (a) help, give (an) authorization'

b. dar *(un) beso, dar *(un) golpe, dar *(un) salto, dar *(una) bofetada

'give *(a) kiss, give *(a) blow, take *(a) jump, give *(a) slap'

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 adjectival modification in the languages with bigger corpora, English (Figure 8) and Spanish (Figure 11), although the percentages are 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 (Figure 9). 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 more frequently modified (Figure 10).

The differences in these tendencies, as well as the effect of bare nominals in the presence or lack of modification in Romance languages confirm a need for a more fine-grained analysis of each LVC selected for this sample.

3.4.2.2. English give-LVCs

As results for the degree of frequency have shown above, high-frequent English *give*-LVCs present a clear tendency towards the modification of the nominal element, followed by medium-frequency and low-frequency to a lesser extent.

Figure 12 presents the distribution of modification patterns (including all subclasses) for English *give*-LVCs per lemma. An examination of each LVC in my sample shows that high-frequency LVCs present the highest percentages of modification in three cases: *give an answer* (82.50%), *give advice* (65.14%), and *give an example* (61%), whereas *give a speech* (37.50%) appears modified to a degree below the mean, which is 53.26% (see Figure 1). Likewise, a similar behaviour can be found in medium-frequency LVCs, where some LVCs are usually modified, such as *give a smile* (70.47%), *give a kiss* (66.50%) and *give a hug* (52.26%), while *give a try* is modified by an adjective in a minority of cases (11.84%). In low-frequency LVCs there is a clear difference between those that are generally modified, such as *give notice* (67.07%), *give a push* (53.49%), *give a sigh* (75%), and the least modified: *give a nod* (34.21%) or *give chase* (1.10%).
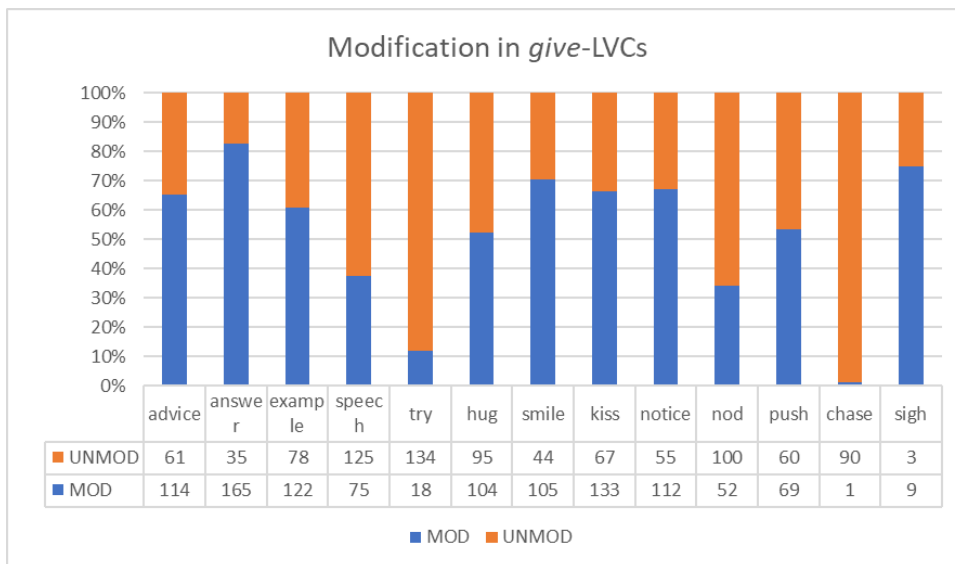
Figure 12. Distribution of modification in English give-*LVCs per lemma*

Table 23 shows the distribution of modification subclasses with a predominance of adjectival modification in all lemmas, except for *give a try* (11.84%) and *give chase* (1.10%), which show very low proportions of adjectival modification, is in line with the general very low distribution of modification in these two LVCs (cf. Figure 12).

Modification	advice	answer	example	speech	try	hug	smile	kiss	notice	nod	push	chase	sigh
Unmod (No)	61 (34.85%)	35 (17.50%)	78 (39%)	125 (62.50%)	134 (88.16%)	95 (47.74%)	44 (29.53%)	67 (33.50%)	55 (32.93%)	100 (65.79%)	60 (46.51%)	90 (98.90%)	3 (25%)
Adj	106 (60.75%)	145 (72.50%)	61 (30.50%)	46 (23%)	18 (11.84%)	95 (47.74%)	97 (65.10%)	117 (88.50%)	42 (25.15%)	38 (25%)	63 (48.84%)	1 (1.10%)	5 (41.67%)
PP	1 (0.57%)	-	36 (18%)	5 (2.50%)	-	-	-	-	8 (4.79%)	7 (4.61%)	-	-	4 (33.33%)
Rel	1 (0.57%)	12 (6%)	2 (1%)	4 (2%)	-	2 (1%)	3 (2.01%)	2 (1%)	3 (1.80%)	2 (1.31%)	5 (3.88%)	-	-
NC	5 (2.85%)	5 (2.50%)	-	15 (7.50%)	-	7 (3.52%)	2 (1.35%)	14 (7%)	9 (5.39%)	2 (1.31%)	-	-	-
Gen	-	-	-	-	-	-	-	-	29 (17.37%)	-	-	-	-
Various	1 (0.57%)	3 (1.50%)	23 (11.50%)	5 (2.50%)	-	-	3 (2.01%)	-	21 (12.57%)	3 (1.98%)	1 (0.77%)	-	-

TOT AL	175	200	200	200	152	199	149	200	167	152	129	91	12

Table 23. Distribution of modification subclasses in *give*-LVCs per lemma

Regarding the rest of modification subclasses, most of them present residual results (3% or below) with some exceptions. For instance, some NVEs select PP as modifiers: *example* (18%), *notice* (4.79%), *nod* (4.61%) and *sigh* (only 4 instances, 33.33%), as illustrated in (39); only one LVC has relative clauses as modifiers, that is the case of *answer* (6%), as in (40); three LVCs combine with noun classifiers: *speech* (7.50%), *kiss* (7%), and *notice* (5.39%), as in (41); and only one LVC presents a considerable proportion of genitive modification, *notice* (17.30%), as in (42).

(39)   a. I will give you *one more example of a smaller state* where the slice of
        the Latino population is smaller   (COCA, 2012)

        b. And do not forget to give your parents *notice of the meeting...* which
        handed out earlier.   (COCA, 2013)

        c. and lift it over your head after you give me *a nod of permission*.
                                                            (COCA, 2012)

        d. he will give *a sigh of relief* and lay down his heavy burden.
                                                            (COCA, 2009)

(40)   In that moment Jesus did not give the people *the answer they were looking
        for.*   (COCA, 2011)

(41)   a. And Cuba Gooding for showing us how to give *our acceptance speech*.
                                                            (COCA, 2013)

        b. I want to give Anita *a goodnight kiss*!   (COCA, 2015)

        c. after you give your official *two-week notice*.   (COCA, 2010)

(42)   also don't rule out the possibility that Bush still might give *6 months' notice*
        that Washington will leave the treaty.   (COCA, 2001)

Going back to the distribution of the adjectival modification in English *give*-LVCs per lemma, it presents an irregular distribution. In the cases of the most modified English *give*-LVCs, an examination of the co-occurrence with certain adjectives fails to show a tendency to collocate, since there is a great variety of adjectivization in all cases. However, some co-occurrences are relevant for the analysis: *give an answer* collocates with *straight* (31/148, 20.95%), as in (43); *give advice* has a tendency to co-occur with

good (12/107, 11.21%), as in (44); and *give a kiss* in singular co-occurs with *big* (27/117, 23.09%), as in (45a), but in plural with *sweet* (4/117, 3.42%), as in (45b).

(43)   I give myself *very good advice* but I very seldom follow it.

(COCA, 2012)

(44)   "Did you kiss him? Just give me *a straight answer*, please."

(COCA, 2015)

(45)   a. Why don't you go over and give her *a big kiss*?   (COCA, 2013)

b. He continues to give her *sweet little kisses.*   (COCA, 2007)

Likewise, the less modified LVCs do not show tendencies to co-occur with certain adjectives. When there are few examples, the adjectives are barely repeated, as in *give a sigh* (46) or *give a try* (47). The only instance of modified nominal in *give chase* is the adjective *good* (48).

(46)   a. She'll look me over, give *a long-suffering sigh*, and announce my BMI and percentage of body fat with alarming accuracy.   (COCA, 2008)

b. Laika would settle onto the floor and give *a heavy sigh.*

(COCA, 2013)

c. Oh, yeah, I often give *a big sigh* every so often.   (COCA, 2012)

d. Mr. Woodhouse could only give *a submissive sigh*   (COCA, 2012)

(47)   a. I'm gonna give things *a real try*.   (COCA, 2017)

b. I would be willing to give it *a second try*.   (COCA, 2012)

c. Some give it *a feeble try*,   (COCA, 2012)

d. he takes a leave of absence, okay, to give it *one last try*,   (COCA, 2011)

(48)   She looks like she'll give *a good chase.*   (COCA, 2001)

In the cases of *give a try* and *give chase* their reluctance to include an adjectival modifier within the nominal element could be taken as a sign of an initial stage of lexicalization. However, the rest of individual English *give*-LVCs show free combination with their NVEs through determination, as well as modification of the NVE. The preference of these LVCs for adjectival modification also conforms to the general tendency of the language in the nominal domain, and only in a minority of cases PPs, relative clauses, noun classifiers and genitives are also found (39)-(42).

3.4.2.3. German geben-LVCs

Figure 13 presents the results of adjectival modification for each German *geben*-LVC. Interestingly, the most modified LVCs are *Beschreibung geben* (70.31%) and *Warnung geben* (36.17%), which are low-frequency LVCs in the corpus; whereas the other two low-frequency LVCs are modified less often: *Kuss geben* (24.55%) and *Verspechen geben* (27.38%). This explains why low-frequent *geben*-LVCs presented a percentage close to that of high-frequency German LVCs in Figure 9. After these, the high-frequency German *geben*-LVCs are modified in more than 31.76% of cases (which is the general mean of German *geben*-LVCs, as seen in Figure 7): *Antwort geben* (35.50%) and *Hinweis geben* (37.95%). In contrast, medium-frequency LVCs show the lowest percentages of adjectival modification: *Rat geben* (27%), *Unterricht geben* (28.45%) and *Erlaubnis geben* (6.67%).

Figure 13. Distribution of modification in German geben-*LVCs per lemma*

Table 24 shows the distribution of modification subclasses with a predominance of adjectival modification in all lemmas, and a residual presence of the other subclasses.

Modification	Hinweis	Antwort	Rat	Unterricht	Erlaubnis	Kuss	Warnung	Beschreibung	Versprechen
Unmod (No)	103 (62.05%)	129 (64.50%)	146 (73%)	38 (77.55%)	70 (93.33%)	83 (75.45%)	30 (63.83%)	19 (29.69%)	61 (72.62%)
Adj	58 (34.94%)	69 (34.50%)	53 (26.50%)	11 (28.45%)	5 (6.67%)	24 (21.82%)	17 (36.17%)	21 (32.81%)	18 (21.43%)
PP	-	-	-	-	-	-	-	3 (4.69%)	-
Rel	4 (2.41%)	2 (1%)	1 (0.50%)	-	-	1 (0,91%)	-	-	5 (5.95%)
NC	-	-	-	-	-	1 (0.91%)	-	-	-
Gen	-	-	-	-	-	-	-	11 (17.19%)	-
Various	1 (0.60%)	-	-	-	-	1 (0.91%)	-	10 (15.62%)	-
TOTAL	166	200	200	49	75	110	47	64	84

Table 24. Distribution of modification subclasses in *geben*-LVCs per lemma

The only exceptions are the low-frequency *geben*-LVCs: *Beschreibung geben* and *Versprechen geben*. In the case of *Beschreibung,* there are instances of PP modification (4.69%), genitive (17.19%), and a combination of adjectives and others (15.62%), as illustrated in (49); while in *Versprechen* the only subclass that is present corresponds to relative clauses (5.59%), as in (50).

(49)   a. Geben Sie *eine Beschreibung von Juan Ryness* an jede Streifeneinheit der Stadt aus,   (DWDS, 2007)

'Give a description of Juan Ryness in every patrol division of the city'

b. «Diese Rede und *diese Beschreibung Europas* hätten sie schon vor zwei Jahren geben müssen», ruft er Merkel zu.   (DWDS, 2012)

'They should have given this speech and this description of Europa two years ago, he called Merkel out'

c. Die restlichen Mitglieder der falschen Denkfabrik haben uns *eine detaillierte Beschreibung von seinen Merkmalen* gegeben.

(DWDS, 2010)

'The rest of the fake Thinktank members gave us a detailed description of their characteristics'

(50)   an dieser Stelle kann ich Ihnen nicht alle Versprechen geben, *die Sie gerne von mir hätten.*   (DWDS, 2006)

'in this position I cannot give you all promises, which you would happily receive from me'

If the co-occurrence with certain adjectives is examined, a high-frequent LVC, *Antwort geben* ('give answer')*,* presents a tendency to co-occur with certain adjectives (51): *klar* ('clear') in 10/69 of modified instances (14.49%) and *richtig* ('correct') in 8/69 (11.50%). This represents a 13.43% (18/134) of all modified instances of high-frequency LVCs.

(51)   a. Auch unser Staatspräsident hat *die klare Antwort gegeben*

(DWDS, 2000)

'Also our state president has given the clear/straight answer'

b. Die Geldpolitik hat *die richtige Antwort gegeben* (DWDS, 2000)

'The monetary policy has given the right answer'

In medium-frequency *geben*-LVCs there is a tendency of *Rat geben* ('give advice') to co-occur with *gut* ('good') and its comparative form *besser* ('better') in 18/53 instances (33.96%); more specifically, 14 instances with *good* (52a) and 4 instances with *besser* (52b).

(52)   a. als Prager Professoren *den guten Rat gaben*.   (DWDS, 2008)
'as the Prague Professors gave the good advice'
b. der die selben Probleme wie Du hatte und kann Dir *besseren Rat geben*.

(DWDS, 2009)

'who has the same problems as you and can give you the best advice'

Also, low-frequency LVCs present certain co-occurring tendencies which should be taken cautiously due to the number of total instances. That is the case of *Beschreibung geben* ('give a description') with *genaue* ('exact') in 7/31 of instances modified by an adjective (22.58%), as in (53), and *Versprechen geben* ('give promise') with *groß* ('big') in 4/18 of modified instances (22.22%), as in (54).

(53)   a. Ich hab dem Typen *eine genaue Beschreibung gegeben*.

(DWDS, 2002)

'I have given the guy an exact description'

(54)   Sie haben *große Versprechen gegeben*,   (DWDS, 2010)

'They have made big promises'

In contrast, there are other *geben*-LVCs which show a wide variety of adjectives modifying the nominal in them, as in *Kuss geben* (55) and *Warnung geben* (56).

(55)   a. *Den schönsten Kuss* gibt der Stehgeiger Jakob seiner geliebten Barbara

(DWDS, 2017)

'The café violinist Jakob gives his believed Barbara the most beautiful kiss'

b. Tanten, die *zu viele feuchte Küsse* geben,   (DWDS, 2004)

'Aunts, who give too many wet kisses'

c. aber haben Sie da Bridget eben *einen kleinen Kuss* gegeben?

(DWDS, 2002)

'but have they even given Bridget a little kiss?'

d. Ich hab' ihr *einen dicken Kuss* gegeben.   (DWDS, 2004)

'I have given her a thick kiss'

(56)   a. *Die zweite Warnung* gab Preis   (DWDS, 2010)

'Preis gave the second warning'

b. Gib ihr *eine letzte Warnung*.   (DWDS, 2005)

'Give her one last warning'

c. Er gibt ihr *eine schriftliche Warnung*.   (DWDS, 2015)

'He gives her a written warning'

To sum up, *geben*-LVCs show a clear tendency to select unmodified NVEs, and those LVCs with more modification show a co-occurrence tendency with certain adjectives.

3.4.2.4. Catalan donar-LVCs

Catalan *donar*-LVCs in my sample have shown a greater tendency to be modified in low-frequency LVCs (see Figure 10). However, these results should be taken cautiously due to the low number of instances extracted for this group of LVCs. Hence, a need for an analysis of each LVC is especially relevant for this language.

Taking into account that the average of modified *donar*-LVCs in Catalan is 28.57% (see Figure 7 above), Figure 14 shows that there is only one LVC with percentages of modification above 50%: *donar definició* (64.71%). Amongst the most modified are *donar resposta* (43.90%), *donar explicació* (41.07%), *donar cop* (40.87%),

donar empenta (40%), *donar ajut/da* (36.84%), *donar conferència* (36.84%), and *donar consell* (36.36%). The rest of the lemmas present percentages which are below the average for the language, as in *donar bes/ada* (20%), *donar pallissa* (16.67%), *donar suport* (11%), *donar bufetada* (8.33%), *donar permís* (4.41%) and *donar volta* (2.47%). Low-frequency LVCs, however, represent a very low number of instances and these percentages must be taken with due precautions. The reason for the low representation of some particular LVCs in the CTILC corpus could be due to the coexistence of this constructions with other LVCs with the more popular LV *fer* ('do/make') (GIEC, 2018), like *donar bes/ada* vs. *fer petó* and *donar/fer conferència*[29].

	supor t	cop	respo sta	volta	explic ació	conse ll	permí s	empe nta	confe rència	ajuda	bufet ada	palliss a	bes/a da	defini ció
UNMOD	178	68	92	79	66	42	65	15	12	12	11	5	8	12
MOD	22	47	72	2	46	24	3	10	7	7	1	1	2	22

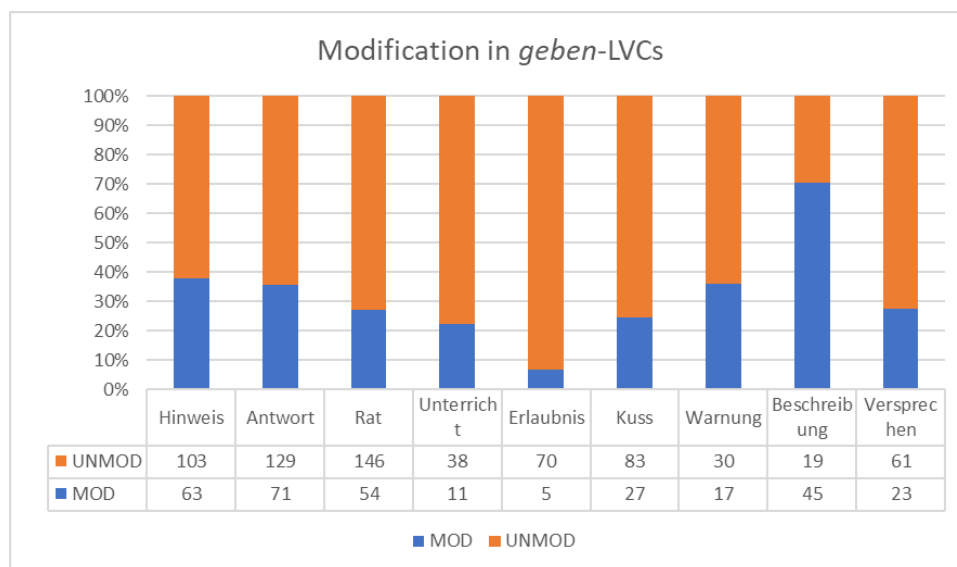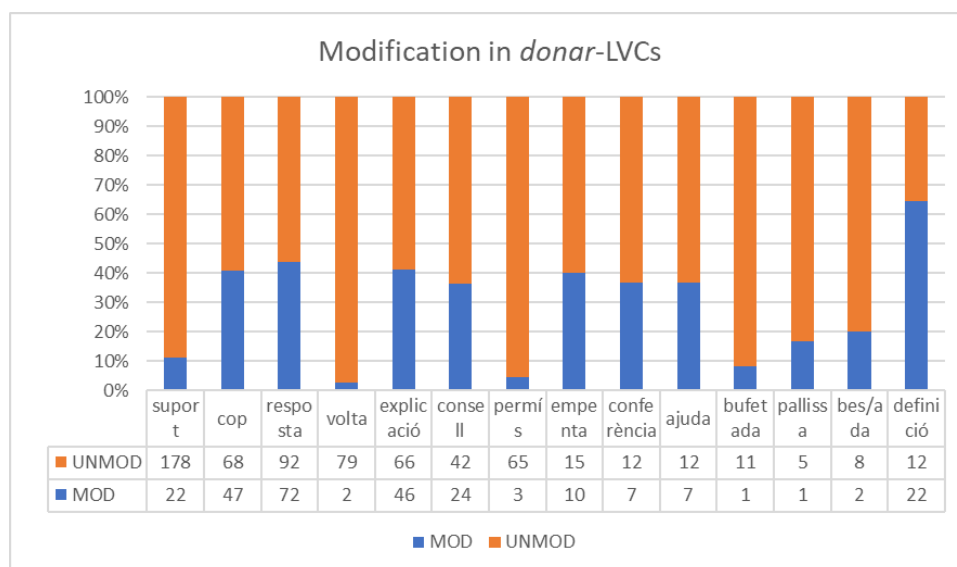*Figure 14. Distribution of modification in Catalan* donar-*LVCs per lemma*

Table 25 shows the distribution of modification subclasses with a predominance of adjectival modification in the majority of lemmas.

---

[29] Other extralinguistic features could also contribute to this, since there is a lower representation of non-Central varieties of Catalan which favour some constructions with *donar* ('give').

Modification	suport	cop	resposta	volta	explicació	consell	permís	empenta	conferència	ajut/da	bufetada	pallissa	bes/ada	definició
Unmod (No)	178 (89%)	68 (59.13%)	92 (56.10%)	79 (97.53%)	66 (58.93%)	42 (63.64%)	65 (95.59%)	15 (60%)	12 (63.16%)	12 (63.16%)	11 (91.67%)	5 (83.33%)	8 (80%)	12 (35.29%)
Adj	20 (10%)	18 (15.65%)	62 (37.80%)	2 (2.47%)	25 (22.32%)	19 (27.94%)	1 (1.47%)	9 (36%)	2 (10.53%)	6 (31.58%)	1 (8.33%)	-	2 (20%)	13 (38.24%)
PP	2 (1%)	27 (23.48%)	2 (1.22%)	-	14 (12.50%)	5 (7.35%)	2 (2.94%)	-	4 (21.05%)	-	-	-	-	5 (14.71%)
Rel	-	1 (0.87%)	8 (4.88%)	-	1 (0.89%)	-	-	1 (4%)	-	-	-	1 (16.67%)	-	-
NC	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Various	-	1 (0.87%)	-	-	6 (5.36%)	-	-	-	1 (5.26%)	1 (5.26%)	-	-	-	4 (11.76%)
TOTAL	200	115	164	81	112	68	68	25	19	19	12	6	10	34

Table 25. Distribution of modification subclasses in *donar*-LVCs per lemma

There are, however, some relevant exceptions, that is, LVCs with prevalence of other subclasses of modification. For instance, PP modification is prevalent in the following lemmas: *donar cop* (15.65% of adjectival modification versus 23.48% of PP modification), and to a lesser extent also *donar conferència* (10.50% adjectival versus 21.05% PP) or *donar* permís (1.47% adjectival versus 2.94% PP), as in (57).

(57)   a. no es varen donar mútuament *un cop de pic*.   (CTILC, 2002)

'they did not hit each other with a pick-axe'

b. Va ser tan ben acollit als Estats Units que després de donar *les conferències de Boston* decidí quedar-s'hi.   (CTILC, 2007)

'He was so welcomed in the United States that after giving the conferences in Boston he decided to stay'

There is also a relevant presence of PP modification in other lemmas: *donar explicació* (12.50%), *donar consell* (7.35%) and *donar definició* (14.71%), as illustrated in (58). In the case of relative clauses, their presence is only relevant in the case of *donar resposta* (4.88%) and *donar empenta* (4%, but only 1 instance), as in (59).

(58)   a. em va donar *una explicació de senzillesa aclaparadora*, no vull tractes,

(CTILC, 2008)

'he gave me an explanation of oppressive simplicity: I don't want any deal'

b. on fins i tot dóna *consells de puericultura*: no a les sopes de pa amb vi i sí a l'alletament matern. (CTILC, 2007)

'where she even gives advice on childcare: no to bread soup with wine and yes to breastfeeding'

c. dóna una *definició d'ideologia* que ens anirà bé per encarar el propòsit que acabem de plantejar. (CTILC, 2015)

'he gives a definition of ideology which will help us face the aim which we just outlined'

(59)   a. i que al mateix temps es donessin *respostes que regulessin la conflictivitat social en el territori.* (CTILC, 2016)
'and at the same time answers would be given which could regulate the social conflict in the country'
b. va trobar una noia qualsevol, capaç de donar-li *l'empenteta que encara li faltava.* (CTILC, 2002)
'he found a random girl, capable of giving him the little push that he still needed'

A closer look at the LVCs with a higher number of instances (*suport* 'support', *cop* 'blow' and *resposta* 'answer') confirms that there are certain restrictions to adjectival modification which are related to the bareness of the nominal element. There is a narrow variety of adjectives modifying bare nominal, as they are mainly combined with relational adjectives (60), and less often with qualitative adjectives (61).

(60)   a. una entitat catòlica que donava *suport mèdic i financer* a les dones.

(CTILC, 2006)

'a catholic institution which gave medical and financial support to women'

b. cosa que permetria *donar respostes educatives* als alumnes menys motivats pels ensenyaments més teòrics   (CTILC, 2000)

'which would allow to give educative answers to the students least motivated by the more theoretical topics'

(61)   a. No tenia forces per donar *respostes més llargues*. (CTILC, 2001)

'(He/She) did not have energy to give longer answers'

b. Els càlculs més fiables realitzats per experts donen *una resposta contundent* als dubtes sobre el perill d'una guerra de destrucció massiva.

(CTILC, 2000)

'The more reliable calculations made by experts give a conclusive answer to the doubts regarding the doubts about the danger of a mass destruction war'

In the case of *donar cops* (lit. 'to hit'), which needs the presence of a determiner in the singular form of the nominal element, the adjectives with which it combines are mostly qualitative. It further tends to co-occur with the adjective *fort* ('strong') in 7 out of 19 occurrences modified by an adjective (36.84%), as in (62).

(62)   a. cal situar la ganiveta sobre el punt escollit i donar *un o dos cops forts* amb una petita maça de fusta.   (CTILC, 2000)

'it is necessary to place the blade on the chosen point and make one or two strong blows with a small wooden mallet'

With regard to low-frequency LVCs, the number of instances is insufficient to detect any tendencies of co-occurrence with certain adjectives.

3.4.2.5. Spanish dar-LVCs

Figure 15 shows that the tendency presented in section devoted to the degree of frequency (previously seen in §3.4.2.1.) is confirmed in the analysis of each LVC for Spanish. The percentage of modified nominals in the medium-frequency group is generally higher than the average for the Spanish *dar*-LVCs (27.69%, see Figure 7 above): *dar cambio* (62.30%), *dar golpe* (41.71%), *dar salto* (35%), and *dar clase* (34%); while *dar consejo* is just below the average (26%). In high-frequency LVCs there is more variation: *dar respuesta* (34%) and *dar paso* (30%) show a relevant tendency to be modified, whereas *dar beso* (25%) and *dar vuelta* (11.5%) are below the average. Finally,

low-frequency is also a varied group: *dar giro* (51.1%) and *dar ayuda* (32%) have a higher presence of modification than *dar bofetada* (14%), *dar instrucción* (19%), *dar autorización* (11.5%) and *dar permiso* (14.5%). In fact, the only two LVCs with a majority of modified instances are *dar cambio* (62.30%) and *dar giro* (51.1%), contrasting with a general tendency for *dar*-LVCs to appear unmodified.
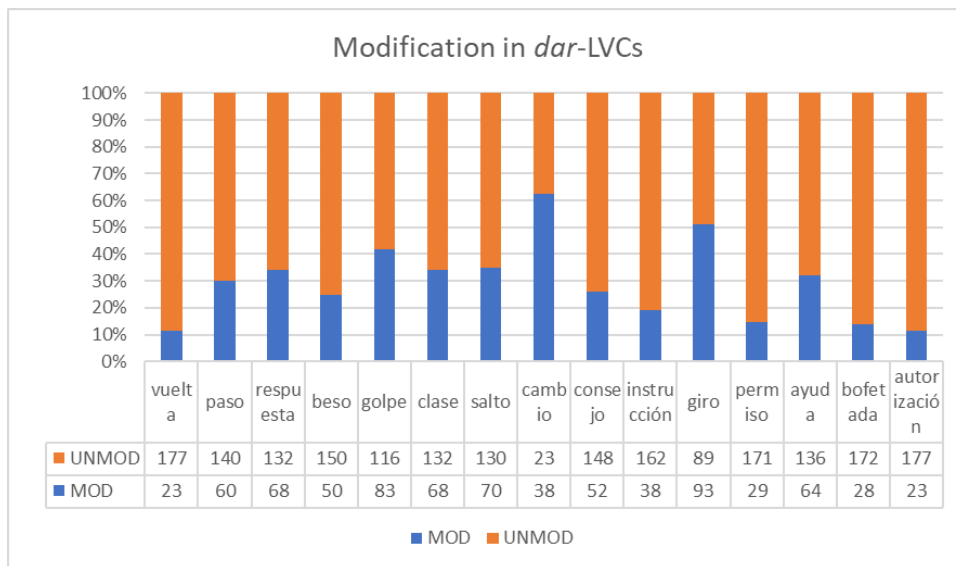
Figure 15. Distribution of modification in Spanish dar-*LVCs per lemma*

Table 26 shows that all lemmas have a preference for adjectival modification.

Modification	vuelta	paso	respuesta	beso	golpe	clase	salto	cambio	consejo	instrucción	giro	permiso	ayuda	bofetada	autorización
Unmod (No)	177 (88.50%)	140 (70%)	132 (66%)	150 (75%)	116 (58.29%)	132 (66%)	130 (65%)	23 (37.70%)	148 (74%)	162 (81%)	89 (48.90%)	171 (85.50%)	136 (68%)	172 (86%)	177 (88.50%)
Adj	19 (9.50%)	53 (26.50%)	58 (29%)	30 (15%)	52 (26.13%)	21 (10.50%)	55 (27.50%)	34 (55.74%)	42 (21%)	28 (14%)	69 (37.91%)	6 (3%)	50 (25%)	26 (13%)	13 (6.50%)
PP	4 (2%)	7 (3.50%)	2 (1%)	17 (8.50%)	28 (14.07%)	47 (23.50%)	13 (6.50%)	2 (3.28%)	6 (3%)	9 (4.5%)	22 (12.09%)	22 (11%)	6 (3%)	-	10 (5%)
Rel	-	-	6 (3%)	2 (1%)	2 (1%)	-	1 (0.5%)	-	3 (1.5%)	-	-	1 (0.5%)	7 (3.5%)	2 (1%)	-
NC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Various	-	-	2 (1%)	1 (0.50%)	1 (0.51%)	-	1 (0.5%)	2 (3.28%)	1 (0.5%)	1 (0.5%)	2 (1.10%)	-	1 (0.5%)	-	-

TO TA L	200	200	20 0	200	199	200	200	61	200	200	182	200	200	20 0	200

Table 26. Distribution of modification subclasses in *dar*-LVCs per lemma

There are two exceptions: *dar clase* (10.57% of adjectival modification versus 23.50% of PP) and *dar permiso* (3% adjectival versus 11% PP), as in (63). Also, there are other lemmas with a relevant PP modification: *dar giro* (12.09%), *dar beso* (8.50%), *dar salto* (6.50%), *dar autorización* (5%), as in (64). The rest of subclasses are, however, residual. There is a similar tendency in the two Romance languages studied in this dissertation.

(63)   a. De la misma forma, Jada Pinkett Smith *da clases de sutileza dramática* en el rol de la esposa de Jonhson,   (CORPES XXI, 2007)

'Likewise, Jada Pinkett Smith gives lessons of dramatic subtlety in the role of Jonhson's wife'

b. Porque, según tengo entendido, *ha dado permiso de obras* para hacer pisos de cinco plantas pegados a las terrazas de los adosados

(CORPES XXI, 2004)

'Because, as far as I know, he gave permission of construction work to build five store buildings next to the semi-detached houses terraces'

(64)   a. Mis hermanas le *dieron un beso de agradecimiento*, y abuela se sonó la nariz varias veces.   (CORPES XXI, 2001)

'My sisters gave her an appreciation kiss, and the grandma blowed her nose several times'

b. *dando saltitos de alegría* gritaban al doctor Remies:

(CORPES XXI, 2002)

'taking joyous jumps they called doctor Remies'

c. Antes de *dar la autorización de salida,* los técnicos hicieron las últimas comprobaciones a la aeronave   (CORPES XXI, 2010)

'Before giving authorization of departure, the technicians made the last checks to the aircraft'

An analysis of the co-occurrence of Spanish *dar*-LVCs with specific adjectives shows that there are tendencies which might have an effect in the percentages of modification in both high and medium-frequency LVCs.

In high-frequency *dar*-LVCs there are two clear examples: *dar una vuelta* ('take a turn') displays a 78.95% of co-ocurrence with the adjective *media* ('half'), with 15 out of 19 instances, as in (65), whereas in *dar un paso* ('take a step') there is a 41.51% of collocation with *primer/o* ('first') with 22 out of 53 instances as in (66).

(65)   a. Amparo *dio media vuelta* y se alejó con paso lento.

                          (CORPES XXI, 2001)

      'Amparo turned and got away slowly'

(66)   Se *da el primer paso* para desarrollar un test de inteligencia universal

                          (CORPES XXI, 2011)

      'The first step is taken to develop a universal intelligence test'

In medium-frequency *dar*-LVCs, a similar tendency is found in *dar un cambio* ('make a change') with two adjectives (67): *radical* ('radical') in 8 out of 36 modified instances (22.22%) and *brusco* ('abrupt') in 6 out of 36 instances (16.67%).

(67)   a. En las últimas tres décadas se *ha dado un cambio radical* sobre la necesidad de   guardar la línea.   (CORPES XXI, 2004)
'In the last three decades, a radical change has been made about the need to watch one's figure'
b. a veces, hay que *dar un cambio brusco* de timón

                          (CORPES XXI, 2006)
'sometimes one must make an abrupt change of direction'

In contrast, such collocations are not found in low-frequency *dar*-LVCs, which present a more varied presence of adjectives.

3.4.3. Discussion

With regard to determination, *give*-LVCs present a tendency to select NVEs introduced by a determiner in all four languages, but this is especially predominant in the two Germanic languages. In English, most *give*-LVCs have an obligatory determiner and the results confirm this tendency with 81.74% of determined NVEs; while German *geben*-LVCs also have a high percentage of determined NVEs (77.99%) with a majority of

lemmas that have an optional determiner. In Spanish, the tendency is to have also *dar*-LVCs with an optional determiner, and the presence of DET is a bit lower (63.27%). Catalan *donar*-LVCs presents the lowest proportion of determined NVEs, just above half of the instances (51.02%), which corresponds to an even distribution between lemmas with obligatory and optional determiner.

Some of the patterns found in the determination of *give*-LVCs can be related to two aspects: the difference between mass nouns and countable nouns (especially in English), and the tendency to collocate between the LV and the NVE (especially when it is a bare singular). In the case of English *give*-LVCs, the only mass noun (*advice*) presents a noticeable proportion of bare singular cases (15%) when compared to the rest of the lemmas. In addition, *give chase* and *give notice* present a majority of instances with bare singular (97.80% and 62.87%, respectively). In the case of German *geben*-LVCs, a similar pattern can be found in *Unterricht geben* ('give lessons'), which presents more instances of bare singulars (63.27%) than determined NVEs (36.73%), unlike the rest of the lemmas. Catalan *donar*-LVCs shows the lowest proportion of determined NVEs, since there are three LVCs with a clear preference for bare singulars, which are among the most frequent in the corpus (that is, *donar suport* 'give support', *donar permís* 'give permission' and *donar ajut/da* 'give help'), while the low-frequency LVCs are most often introduced by a determiner. These results are also combined with low percentages of modification in bare nominals, which could point to a lexicalization stage of some LVCs (see the discussion below). Finally, in Spanish *dar*-LVCs there is more variation in the preference for bare singulars in some cases (i.e. *dar permiso* 'give permission') but bare plurals in others (i.e., *dar clases* 'give lessons' and *dar instrucciones* 'give instructions').

The analysis of these lemmas regarding modification patterns points towards signs of lexicalization of these LVCs, although the boundaries for lexicalization are unclear (Contreras & Suñer, 2004). The fact that some LVCs present signs of lexicalization (through fixed constructions that disfavour both determination and modification), while others present a free combination with the NVE, is in line with previous studies on Spanish LVCs which conclude that there is a gradual process of lexicalization from LVCs in an incipient stage to a more advanced stage (Bustos Plaza, 2005; Camacho Coma, 2020). These studies attribute it to their phraseological nature, whereby the process of lexicalization concludes when they reach the status of idioms (i.e. *tomar el pelo* 'to fool', Camacho Cano, 2020: 36; *dar la lata* 'to hassle', Bustos Plaza, 2005). In this dissertation,

115

however, the focus is on LVCs without an idiomatic reading and have not considered idiomatic LVCs. Even in cases where a stronger fixation can be detected (i.e. *give chase, give notice*), the meaning is not idiomatic, and it is provided by the NVE: the event is to *chase* and to *notice,* respectively.

I follow Sanromán Vilas (2015) who states that most irregularities of LVCs can be explained through the morphosyntactic properties of the NVE and the LVC as a whole, but very little is related to structural fixation. In fact, most lemmas in all four languages exhibit a free, unrestricted combination with the NVE. When limitations are found, such as the restriction on bare nouns or the tendency to accept PP modification, these can be related to general tendencies of the languages involved. This is in line with the hypothesis that LVCs behave as traditional verbo-nominal constructions, as proposed in Bruening (2016).

In a general correlation between determination and modification, the two Romance languages analysed combine more often with bare nominals but these are less often modified, while the two Germanic languages select bare nominals as NVE in *give*-LVCs, but these are modified in a larger proportion.

The main purpose of this fine-grained corpus study was to check how the frequency of modification of the nominal element in LVCs compares between Germanic languages (English and German) and Romance languages (Catalan and Spanish). The results show that modification is only general in English, while it is less frequent in German, Catalan and Spanish to different degrees.

It is important to bear in mind that the results show very little prominence of modification subclasses other than adjectival modification in the four languages. This is also in line with previous research by Levin & 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 my 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 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*, *conferència* and *permís* in Catalan; *clase* and *permiso* 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 this study are slightly lower than the findings by Bonial & Pollard (2020) 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 adjectival modification, as well as quantifiers, determiners and relative clauses.

To my knowledge, all previous studies have only included data from Germanic languages, so the results for Catalan *donar*-LVCs and Spanish *dar*-LVCs are novel and cannot be compared to previous findings. In the case of the two Romance languages, the results on adjectival modification are more similar within the 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. While high-frequency LVCs in English *give*-LVCs and Spanish *dar*-LVCs are more often modified than medium and lower frequency LVCs (see Figures 2 and 5), the effect is less clear in the case of German *geben*-LVCs which presents more modification both in high-frequency and low-frequency than in the medium-frequency group (see Figure 3). Conversely, Catalan shows a clear preference for the modification in low-frequency *donar*-LVCs. The latter can be explained by a

higher presence of nominals with a determiner in this group in contrast to high-frequency *donar*-LVCs in the present sample, which include more bare nominals (see Figure 4 above). The difficulties of modification of bare nominals in languages like Spanish and Catalan (Espinal & McNally, 2011) may be the origin of this mismatch in the results – although this cannot be the only predictor, as the results for Spanish *dar*-LVCs present a more flexible modification. In any event, 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.

Previous studies on the use of LVCs versus their synthetic counterparts (Storrer, 2007; Sanromán Vilas, 2009; Bratánková, 2013) point towards the tendency of frequent LVCs to develop a lexicalized meaning, which would favour their use over that of a synthetic verb. In this study, some of the high-frequency LVCs in the four languages show some tendencies to co-occur with certain adjectives, but this is not generalized. This trend is most frequently found in Romance languages, especially in Spanish *dar media vuelta* 'to turn around' (78.95%) and *dar primer/os paso/s* 'take first steps' (41.51%), but also in Catalan *donar cop/s fort/s* 'give strong blows' (38.89%); whereas it is found to a lesser extent in Germanic languages: *gut/besser Rat geben* 'give good/better advice' (33.96%) and *give straight answer* (21.38%). In contrast, a majority of particular 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.

3.5. Quantitative results: the LV *take*

The light verb *take* has also been widely studied as one of the most frequently used verbs in the English language, according to Huddleston & Pullum (2002). In German, the LV *nehmen* appears in the tenth position of the most frequent verbs of the language (Bruker, 2011), and it has been studied in several studies as one of the most paradigmatic LVs of the German language (Levin & Ström Herold, 2015; Fleischhauer & Gamerschlag, 2019). In comparison, the light verbs *tomar*, in Spanish, and *prendre*, in Catalan, seem to be not so frequent. Although they are still encountered under the lists of the main LVs (Alonso-Ramos, 2004; Sanromán Vilas, 2017; GIEC, 2016: 17.2.2), their combinatorial possibilities with nominal elements are more restricted than those of the other two verbs under study in this dissertation: *dar/donar* ('give') and *fer/hacer* ('make').

Table 27 (repeated here for convenience) presents LVCs with the verb *take* in the four languages under study. The selection of LVCs is based on previous studies: for English, Garcia-Vega & Machonis (2014); for German, Rug & Tomaszewski (1993: 255-259); for Spanish, Alonso-Ramos (2004: 325-26); for Catalan, *GIEC* (17.2.2). The selection is further based on frequency lists within the different selected corpora.

	High frequency	Medium frequency	Low frequency
Take (English)	step, advantage, action, picture, breath	risk, walk, shower, bite, nap, bath	hike, drink, stroll, swim
Nehmen (German)	*Abschied* ('farewell'), *Rücksicht* ('consideration'), *Einfluss* ('influence'), *Platz* ('place/seat'), *Abstand* ('distance')	*Einsicht* ('insight'), *Anteil* ('interest')	*Anstoß* ('ofence')
Prendre (Catalan)	*decisió* ('decision'), *consciència* ('consciousness'), *nota* ('note')	*possessió* ('possession'), *iniciativa* ('iniciative'), *partit* ('party'), *precaució* ('precaution'), *molèstia* ('bother')	*distància* ('distance'), *determinació* ('determination'), *bany* ('bath'), *resolució* ('resolution'), *impuls* ('boost'), *represàlia* ('reprisal')
Tomar (Spanish)	*decisión* ('decision'), *nota* ('note'), *fotografía* ('picture'), *conciencia* ('consciousness')	*iniciativa* ('iniciative'), *precaución* ('precaution'), *distancia* ('distance')	*determinación* ('determination'), *descanso* ('rest'), *baño* ('bath'), *acuerdo* ('agreement'), *represalia* ('reprisal'), *atajo* ('shortcut')

Table 27. *Take*-LVCs analysed in the four languages under study

The final database of *take*-LVCs instances consists of 6,521 tokens of LVC with some differences in the distribution among the four languages: Spanish and English provide the higher number of examples above 2,000 each, German above 1,000, but only

around 500 for Catalan (Table 28). Once again, these divergences are consistent with the differences presented in the size of the corpora (see section 3.2.3).

	English	German	Catalan	Spanish	TOTAL LV
Take | Nehmen | Prendre |Tomar	2,530	1,120	513	2,358	6,521

Table 28. Distribution of the number of LVCs of the database for each language

Corpus data allow to establish certain patterns as well as compare between and within language families. The upcoming subsections will focus on both determination and modification in *take*-LVCs.

3.5.1. Determination in *take*-LVCs

The results on determination in *take*-LVCs show a majority of determination in English *take*-LVCs (74.43%) and in Spanish *tomar*-LVCs (56.78%), while Catalan *prendre*-LVCs (34.70%) and German *nehmen*-LVCs (12.86%) have significantly lower percentages (see Figure 16).

Determination in *take*-LVCs	English	German	Catalan	Spanish
BPl	463	6	67	510
BS	184	970	268	509
DET	1883	144	178	1339
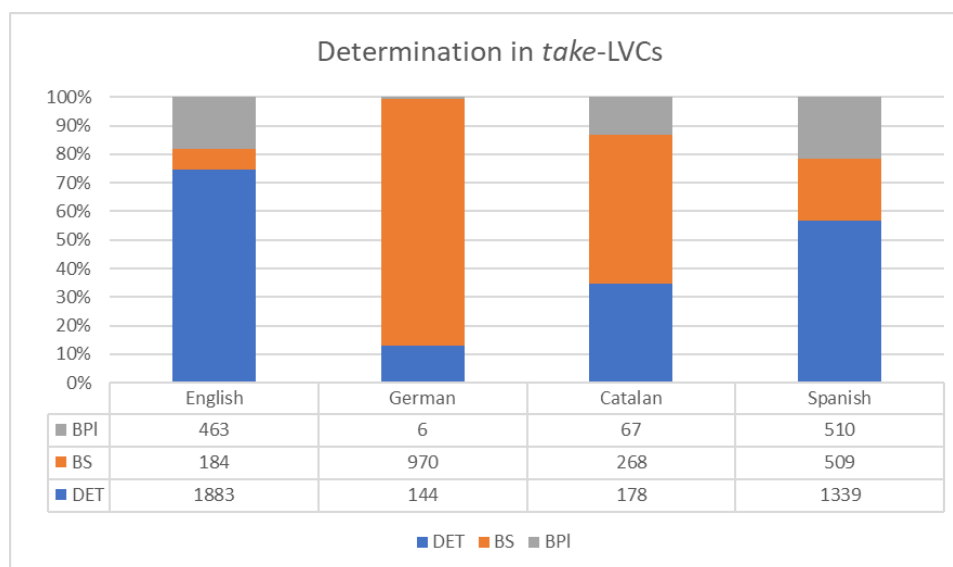
Figure 16. Distribution of the determination in take*-LVCs*

With regards to bare nominals, English presents bare plurals (18.30%) more often than bare singulars (7.27%), which contrasts with the results in *give*-LVCs that preferred bare singulars due to the prevalence of certain LVCs (i.e. *give notice* and *give chase*). In contrast, German results for bare nominals are the opposite: bare singulars are much more frequent (86.60%) than bare plurals (0.54%), contrasting with the tendencies shown for *geben*-LVCs which preferred bare plurals. A similar trend is found in Catalan *prendre*-

LVCs which presents a majority of instances of bare singulars (52.24%), and a lower presence of bare plurals (13.06%). Finally, the results for Spanish bare nominals are quite even: 21.59% of bare singulars and 21.63% of bare plurals. In this case, the distribution is statistically significant when all languages are compared ($\chi2(6)=2601.15$, $p=0$, Cramer's V=0.447), and also when the languages families are compared: Germanic languages ($\chi2(2)= 2266.05$, $p=0$, Cramer's V = 0.788, large effect), and Romance languages ($\chi2(2)= 200.6$, $p<.0001$, Cramer's V=0.264, medium effect).

The distribution of obligatory and optional determiners in *take*-LVCs is especially relevant in the case of German, which includes only LVCs with an optional determiner and could have influenced the results of such low determination percentages for *nehmen*-LVCs. In contrast, English *take*-LVCs tend to select nominals with an obligatory determiner. The two Romance languages show that both Cat. *prendre*-LVCs and Sp. *tomar*-LVCs combine most commonly with optionally determined nominals, and only in 2 and 3 cases (respectively) the determiner is obligatory. The fact that German, Catalan and Spanish combine more often than English with bare nominals (as seen in Figure 16) reflects to a certain extent the distribution of mandatory vs. optional determiners (Table 29).

	Optional DET	Obligatory DET
take	Advantage, action, breath, risk	Step, picture, walk, bite, nap, bath, shower, hike, drink, stroll, swim
nehmen	Abschied ('farewell'), Abstand ('distance'), Rücksicht ('consideration'), Einsicht ('insight'), Einfluss ('influence'), Anteil ('interest'), Platz ('place/seat'), Anstoß ('ofence')	
prendre	Impuls ('boost'), bany ('bath'), represàlia ('reprisal'), resolució ('resolution'), consciència ('consciousness'), nota ('note'), possessió ('possession'), partit ('party'), iniciativa ('initiative'), precaució ('precaution'), distància ('distance'), determinació ('determination')	Decisió ('decisión'), molèstia ('bother')
tomar	Represalia ('reprisal'), nota ('note'), conciencia ('consciousness'), iniciativa ('initiative'), precaución ('precaution'), acuerdo ('agreement'), distancia ('distance'), determinación ('determination'), descanso ('rest'), baño ('bath')	Decisión ('decision'), fotografía ('picture'), atajo ('shortcut')

Table 29. Distribution of obligation and optionality of a determiner in *take*-LVCs

The distribution in Table 29 is aligned with the different tendencies of both language families in the nominal domain, with Romance languages being more flexible in accepting bare nominals in this position than English (Chierchia, 1998; Longobardi, 2001). However, German *nehmen*-LVCs stand out here as having a different tendency from what is expected in Germanic languages regarding bare nominals in object position, as well as in comparison with the two other verbs analysed in this dissertation (*geben* 'give' and *machen* 'make').

Figure 17 shows that English *take*-LVCs follow the tendency presented in Figure 16 with a very general prevalence of determination. The only two exceptions (*take advantage* and *take action*) have more instances of bare singulars than bare plurals (22.61% of bare singulars in *take action*), and even more than determined noun phrases in the case of *take advantage* (67% of bare singulars).
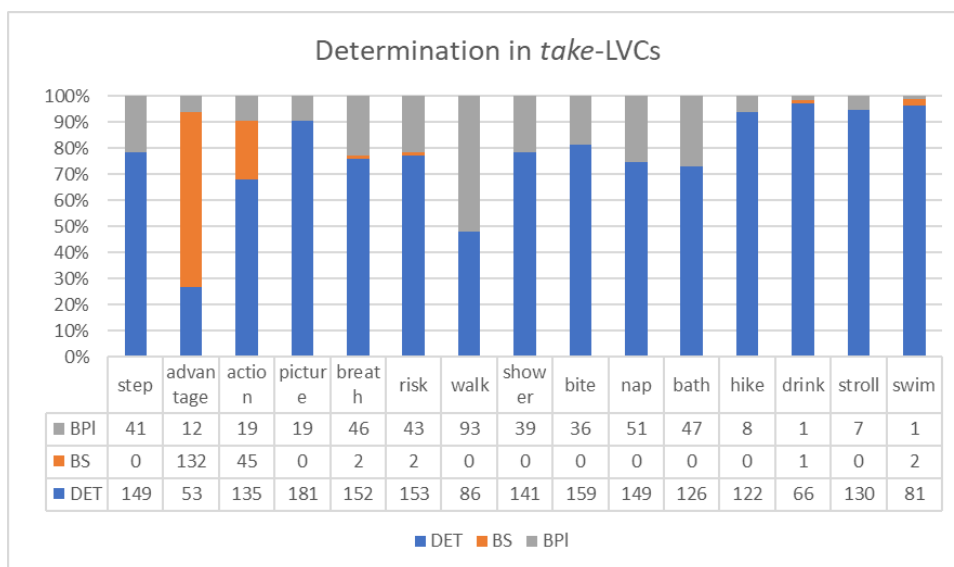
Figure 17. Distribution of the determination in English take*-LVCs per lemma*

The data from these two LVCs exceptions (*take advantage* and *take action,* as in 68), which contrast with the general tendencies found in the rest of lemmas, might point towards a lexicalization of these two LVCs, since they do not display other grammatical factors, such as being mass nouns. This will be further discussed in the upcoming sections with regard to modification.

(68)   a. We are now engaged in a strategic plan process which seeks *to take full advantage* of our location.   (COCA, 2012)

          b. we're giving them all of this information so that they could *take action*, the action that they deemed appropriate.   (COCA, 2017)

Figure 18 shows that German *nehmen*-LVCs follow the basic tendency presented in Figure 16 with a very general prevalence of bare singulars in all lemmas. More specifically, the instances of bare plurals (but also plurals introduced by a determiner) are residual in the selection of these nominals by the LV *nehmen,* which contrasts with the other two verbs and singles this verb out.
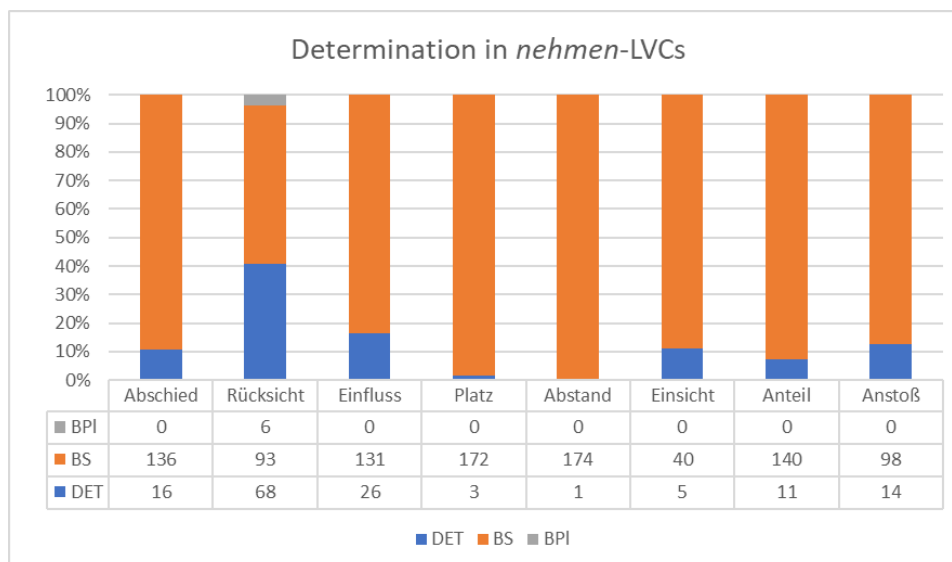
Figure 18. Distribution of the determination in German nehmen-*LVCs per lemma*

Figure 19 shows that the tendency of Catalan *prendre*-LVCs to coappear with bare nominals (in Figure 16) is not a general trend. Although there are some lemmas with a clear predominance of bare nominals (especially bare singulars), other LVCs have a majority of NVEs introduced by a determiner. These are mostly lemmas which have fewer corpus instances and, thus, cannot have a big weight on the general results.
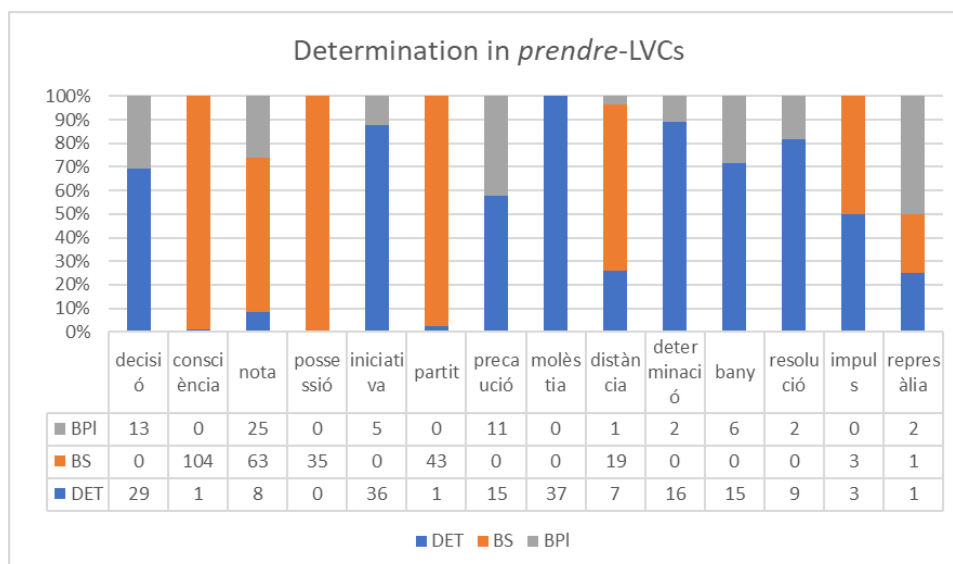
Figure 19. Distribution of the determination in Catalan prendre-*LVCs per lemma*

Apart from *prendre decisió* 'take a decision', which has an obligatory determiner and only accepts bare plurals (69a,b), as well as *prendre iniciativa* 'take initiative', which shows to have lexicalized the definite article (69c), the rest of most frequent LVCs have a clear prevalence of bare singulars: 99.05% in *prendre consciència* 'take awareness',

100% in *prendre possessió* 'take possession', as well as 97.73% in *prendre partit* 'take sides', as illustrated in (70).

(69)   a. Havia *pres una decisió*; no sabia si l'encertava o no, però no hi havia alternativa. (CTILC, 2011)

'He had made a decision; he didn't know if it was the correct one, but there was no alternative'

b. amb un cert aire de persona avesada *a prendre decisions* al llarg del dia. (CTILC, 2013)

'with a certain look of a person used to making decisions throughout the day'

c. Aquell mateix dia, el comandant Fabre *va prendre la iniciativa* d'ajudar Boronat. (CTILC, 2008)

'That same day major Fabre took the initiative of helping Boronat'

(70)   a. Els joves podran *prendre consciència* de la realitat i s'hi revoltaran. (CTILC, 2013)

'The youngsters will be able to be aware of reality and will revolt against it'

b. Rosa Otunbàieva va *prendre possessió* ahir del càrrec com a cap d'Estat. (CTILC, 2010)

'Rosa Otunbaieva yesterday took possession of the position as head of State'

c. Els maçons de París estaven *prenent partit* per la Comuna. (CTILC, 2012)

'Parisian masons were taking sides with the Commune'

Figure 20 shows that tendencies are lemma-dependent in the case of Spanish *tomar*-LVCs. The majority of LVCs present a majority of instances introduced by a determiner. At the same time, some LVCs combine mostly with bare nominals. More specifically, three LVCs combine with bare singulars preferably: *tomar nota* (61.50%), *tomar conciencia* (99.50%) and *tomar distancia* (78.50%); while two LVCs prefer bare plurals: *tomar fotografías* (51.05%) and *tomar represalias* (82.28%).
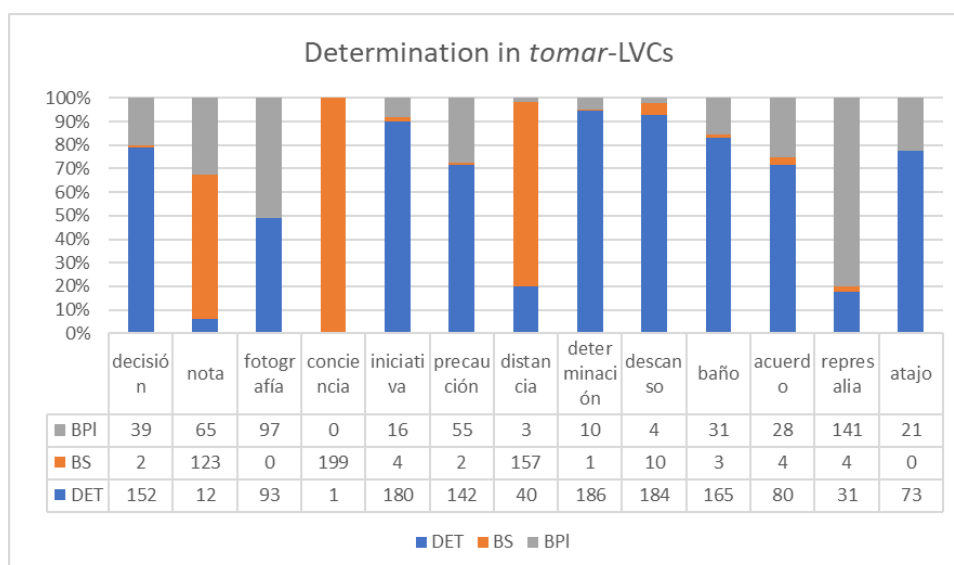
Figure 20. Distribution of the determination in Spanish tomar-*LVCs per lemma*

When determination and modification are compared in bare nominals, modification is more restricted in languages with more presence of bare nominals (German, Catalan and Spanish, as shown above in Figure 16) than in English. Figure 21 shows that English bare nominals are modified in 91.30% of bare singulars and 71.71% of bare plurals, the differences between them being statistically significant ($\chi2$ (1)=28.8, p<.0001, Cramer's V=0.211, small effect). Conversely, Spanish presents much lower percentages, but with little difference depending on number: 19.06% with bare singulars vs. 20.39% with bare plurals. Likewise, Catalan proportions are just slightly above these numbers: 25% with bare singulars vs. 22.39% with bare plurals. The differences for these languages are, thus, not statistically significant ($\chi2$ (1)=0.29, p=0.59, Cramer's V=0.017, negligible effect; and $\chi2$ (1)=0.2, p=0.66, Cramer's V = 0.024, negligible effect, respectively). Finally, German *nehmen*-LVCs have a completely opposite behaviour: only 5.5% of bare singulars are modified, which are the most common, while 66.67% of bare plurals appear with modification, which represents only 4 instances due to the low number of bare plurals in general (a total of 6). In this case, the chi-squared test could not be run because some expected frequencies are too small.
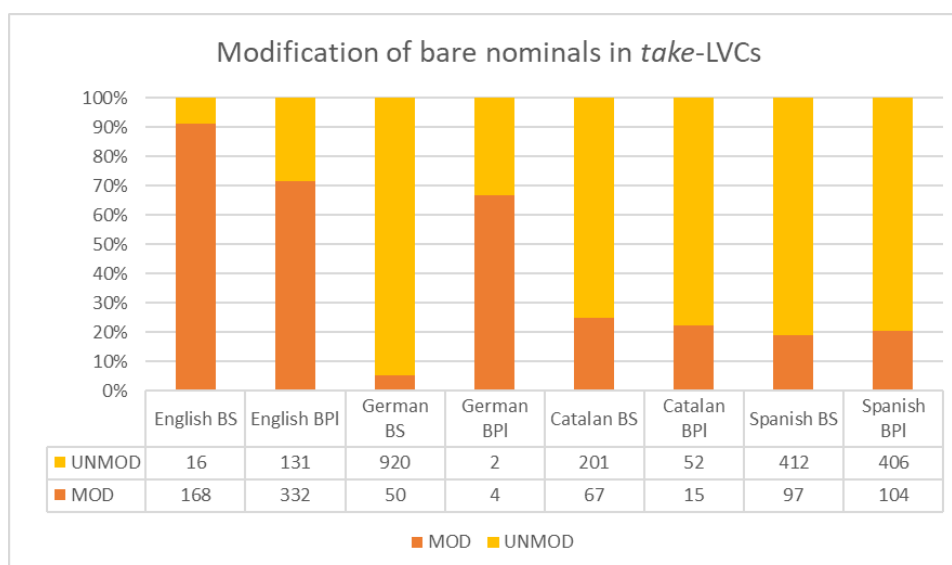
Figure 21. Distribution of the modification of bare nominals in take-*LVCs*

After establishing this first correlation between determination and modification, a description of modification in *take*-LVCs in general is needed. The next section will present an analysis of the different subclasses of modification included in this study, with a special focus on adjectival modification due to its relevance in proportion to its frequency.

3.5.2. Modification in *take*-LVCs

The results on adjectival modification in *take*-LVCs show that the most frequently modified nominal elements within these constructions are found in English (64.74%) contrasting with the results in the rest of the languages: 30.62% in Spanish, 30.02% in Catalan and only 5.71% in German (Figure 22). This distribution is statistically significant when all languages are compared ($\chi 2(3)=1306.31$, p<.0001, Cramer's V=0.448, large effect), but only within one of the language families. When the two Germanic languages are compared, statistical tests prove a significant difference ($\chi 2(1)=1086.95$, p<.0001, Cramer's V=0.546, large effect). Conversely, as the almost equal percentual distribution has already indicated, the difference is not statistically significant for the two Romance languages ($\chi 2(1)=0.07$, p=0.78, Cramer's V=0.005, negligible effect).

127
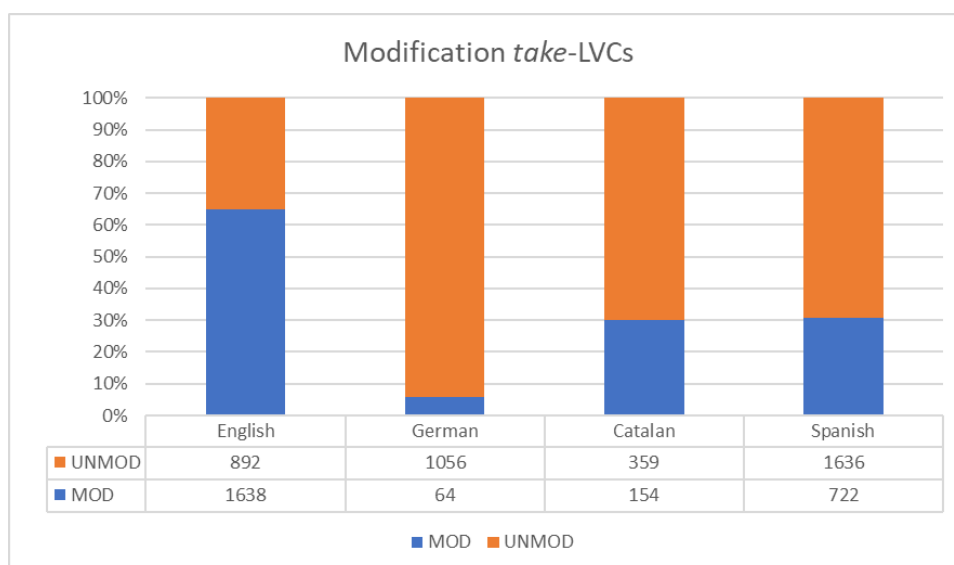
Figure 22. Distribution of modification with take-*LVCs*

The distribution of modification subclasses in the four languages has shown a clear difference between the two Germanic languages, with preference for adjectival modification, and the two Romance languages, with a combination of PP modification and adjectival modification (Table 30).

Modification	English	German	Catalan	Spanish
Unmod (No)	892 (35.26%)	1056 (94.29%)	359 (69.98%)	1636 (69.38%)
Adj	1211 (47.86%)	64 (5.71%)	43 (8.38%)	317 (13.44%)
PP	90 (3.56%)	-	102 (19.88%)	324 (13.74%)
Rel	29 (1.15%)	-	6 (1.17%)	46 (1.95%)
NC	166 (6.56%)	-	-	-
Gen	7 (0.28%)	-	-	-
Various	135 (5.33%)	-	3 (0.59%)	35 (1.49%)
TOTAL	2530	1120	513	2358

Table 30. Distribution of modification subclasses in *take*-LVCs

More specifically, Table 30 shows that adjectival modification is the most common in English (47.86%) in comparison with the rest of subclasses (16.88%), from which only two kinds stand out: noun classifiers (6.56%), as in (71), and a combination of adjective and other kinds of modification (5.33%).

(71)   a. and the winner will take *a one-game advantag*e in the all-time series

            between the two clubs.   (COCA, 2012)

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study

b. How many people now want you to take *their wedding picture*?

(COCA, 2018)

In contrast, German *nehmen*-LVCs only present adjectival modification (5.71%), as in (72), but no instances of any other kind of modification. The most common modification in the two Romance languages is the PP modification: 19.88% in Catalan *donar*-LVCs (versus 8.38% of adjectival modification), as in (73), and 13.74% in Spanish *dar*-LVCs (versus 13.44% of adjectival modification), as in (74).

(72)   Und darauf muss ein Präsident vor einer Entscheidung immer noch *größere Rücksicht* nehmen,   (DWDS, 2002)

‘And a President must take this always into great consideration when facing a decision’

(73)   a. La Conferència va prendre *nota d'una altra Declaració* sobre la subsidiarietat adoptada pels governs d'Alemanya, Àustria i Bèlgica,

(CTILC, 2005)

‘The Conference took note of another Declaration about the subsidiarity adopted by the governments in Germany, Austria and Belgium’

b. fèiem excursions i ens portaven a la platja a prendre *banys de sol*. (CTILC, 2011)

‘[we] did excursions and they took us to the beach to sunbathe’

(74)   a. El Gobierno Nacional tomó *la determinación de suspender las importaciones* de maíz blanco que se habían previsto para finales de 2002.

(CORPES XXI, 2002)

‘The national government took the determination of cancelling the white corn importation which was planned for the end of 2022’

b. Si quiere hasta puede recostarse y tomar un *descanso de diez minutos*.

(CORPES XXI, 2003)

‘If you want, you can even lie down and take a 10-minute break’

The results point towards a difference in the tendency to appear modified between English and the rest of the languages. These divergences will be analysed through a closer examination on the degree of frequency of LVCs (§3.5.2.1) as well as each individual LVC in every language (§3.5.2.2, §3.5.2.3, §3.5.2.4).

3.5.2.1. Degrees of frequency

A closer look at the correlation between the degrees of frequency of LVCs in the corpora and the presence or lack of modification is necessary to determine whether they could have influenced the overall results.

Figure 23 shows that English medium frequency *take*-LVCs appear modified in the majority of cases (71.64%), as well as highly frequent (68.26%), while low-frequency LVCs are modified in only 37.95% of the instances. The differences are statistically significant between these three groups ($\chi2(2)=160.6$, p<.0001, Cramer's V=0.252, medium effect). Thus, the results for English point to a correlation between degree of frequency and modification, since medium- and high-frequent LVCs present much higher results than low-frequent LVCs in modification.
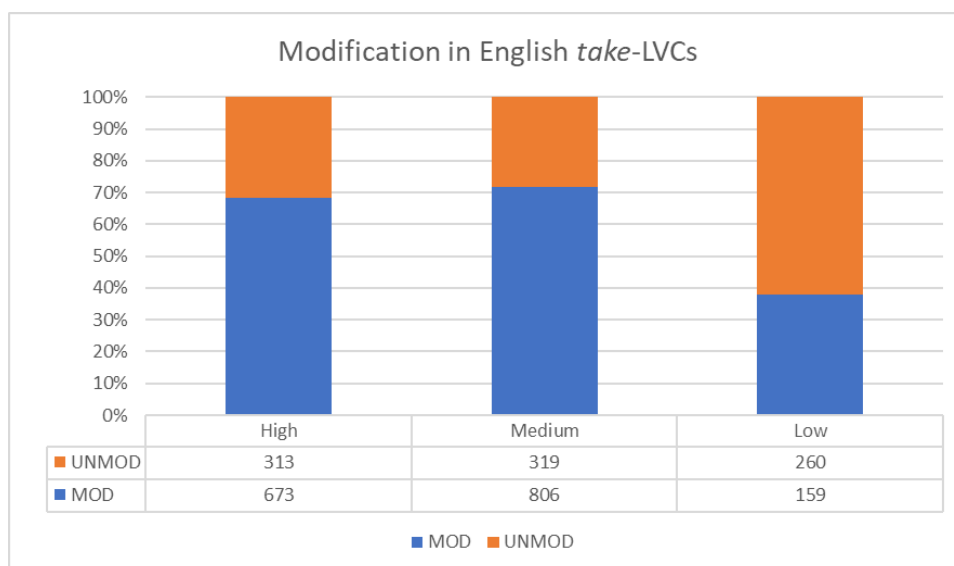
Figure 23. Distribution of modification in English take-*LVCs according to the degree of frequency (high, medium or low)*

Table 31 shows the distribution of modification subclasses, where the adjectival modification is predominant in the three groups: high (45.54%), medium (57.69%), and in low-frequency LVCs (26.97%).

Modification	High	Medium	Low
Unmod (No)	313 (31.74%)	319 (28.36%)	260 (60.05%)
Adj	449 (45.54%)	649 (57.69%)	113 (26.97%)
PP	75 (7.61%)	12 (1.07%)	3 (0.72%)
Rel	13 (1.32%)	14 (1.24%)	2 (0.48%)
NC	19 (1.93%)	110 (9.78%)	37 (8.83%)
Gen	4 (0.40%)	2 (0.17%)	1 (0.24%)
Various	113 (11.46%)	19 (1.69%)	3 (0.71%)
TOTAL	986	1125	419

Table 31. Distribution of modification subclasses in *take*-LVCs according to the degree of frequency

The rest of the modification subclasses are residual (3% or below) except for PP in high-frequency LVCs (7.61%), as in (75), and noun classifiers in both medium-frequency (9.78%) and low-frequency (8.83%), as in (76). The combination of various subclasses (especially Adj + PP) is also considerable in the high-frequency group (11.46%), as illustrated in (77).

(75)   a. Worksites, commercial centres, out of doors holidays and warehouses take *the advantage of walkie talkie radios* to speak with one another.

(COCA, 2012)

b. The man who delivered the knockout blow then returns to take *a cellphone picture of the man*, who is still out cold. (COCA, 2018)

(76)   a. I have quiet time while the baby and toddler take *an afternoon nap.*

(COCA, 2012)

b. I thought we could take *a moonlit stroll* through the marshlands of Oshionian Prime.   (COCA, 2000)

(77)   a. If this sounds too daunting, just take a deep breath and *take the first steps of performing the look-ups* and sending the QWR letter.   (COCA, 2012)

b. Outside I take *a deep breath of somewhat fresher air* and gaze around.

(COCA, 2013)

In a similar line, Figure 24 shows that the German medium-frequent LVCs are much more often modified (11.22%) than the other two groups: 4.96% in high-frequent LVCs and 1.02% in low-frequent LVCs, which is statistically significant ($\chi2(2)=15.92$, p=0.0003, Cramer's V=0.119, small effect). These results align with the English results in

that medium-frequency LVCs are the ones that are most often modified, even though the percentages are much lower compared to those of English *take*-LVCs in the three groups.
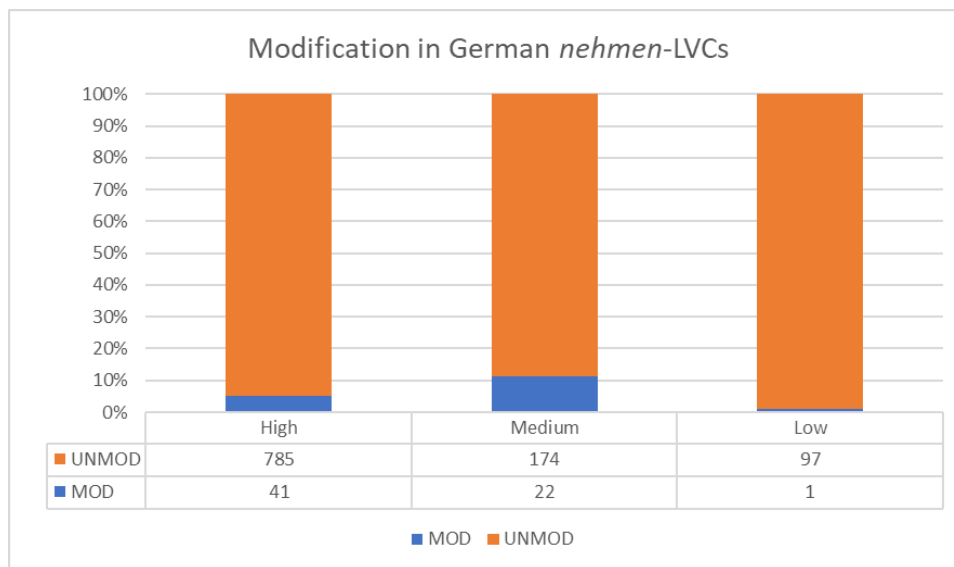
Figure 24. Distribution of modification in German nehmen-*LVCs according to the degree of frequency (high, medium or low)*

There is no presence of additional modification subclasses other than adjectival modification. In fact, the PPs found with a few nouns belong to the verbal paradigm rather than to the nominal modification: *Abschied nehmen von* (78)*, Abstand nehmen von* (79) or *Rücksicht nehmen auf* (80). In all these cases, the preposition is obligatorily selecting an argument of the whole construction rather than the modification of the noun, as their synthetic verbal counterparts also require the preposition (e.g. *sich von jdm. verabschieden* 'to say goodbye to', and *von etwas abstehen* 'to be at a distance from something').

(78)   a. Müssen wir deshalb *Abschied nehmen von* unserem Vertrauen in die Zweckmäßigkeit der Marktwirtschaft?   (DWDS, 2000)

'Therefore, should we take leave from our trust in the usefulness of Capitalism?'

b. Die Familie und die SPD werden gemeinsam von ihm *Abschied nehmen*.

(DWDS, 2004)

'The family and SPD altogether will part from him'

(79)   a. oder sollte man davon eher *Abstand nehmen*?   (DWDS, 2013)

'or should one take distance from it?'

b. Niemand kann vollständig *Abstand nehmen* von seinen Sensibilitäten, seiner persönlichen Herkunft, Kultur und seinem Umfeld. (DWDS, 2007)

'Nobody can completely take distance from his sensibilities, his personal heritage, culture and his environment'

(80)   a. Wie dieser Junge schon mit sechs auf seinen Vater *Rücksicht nehmen* mußte   (DWDS, 2002)

'How this youngster already with 6 years old had to take their father into consideration'

b. und muss auf den Mieter *Rücksicht nehmen*.   (DWDS, 2003)

'and he/she must take the tenant into consideration'

In contrast, Figure 25 shows that the distribution of modification in Catalan *prendre*-LVCs presents a different tendency from Germanic languages as it favours modification in medium-frequency LVCs. The tendency is very similar in the three groups of Catalan *prendre*-LVCs: 31.29% in high-frequency LVCs, 28.86% in medium, and 27.91% in low-frequency. As expected in the face of these numbers, differences are not statistically significant ($\chi2(2)=0.49$, $p=0.78$, Cramer's V=0.031, negligible effect).
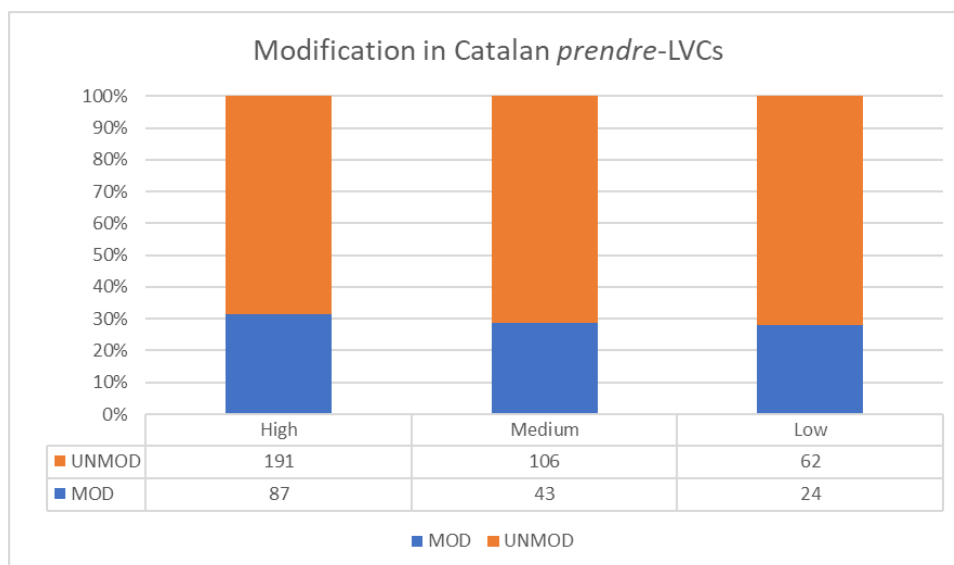
Figure 25. Distribution of modification in Catalan prendre-*LVCs according to the degree of frequency (high, medium or low)*

Table 32 shows that adjectival modification is the preferred subclass of modification only in low-frequency *prendre*-LVCs (17.44% versus 9.30% PP modification); while high-frequency and medium-frequency favour PP modification: 22.30% and 21.48%, respectively. This is illustrated in (81).

Modification	High	Medium	Low
Unmod (No)	191 (68.71%)	106 (71.14%)	62 (72.09%)
Adj	19 (6.83%)	9 (6.04%)	15 (17.44%)
PP	62 (22.30%)	32 (21.48%)	8 (9.30%)
Rel	3 (1.08%)	2 (1.34%)	1 (1.17%)
NC	-	-	-
Gen	-	-	-
Various	3 (1.08%)	-	-
TOTAL	278	149	86

Table 32. Distribution of modification subclasses in *prendre*-LVCs according to the degree of frequency

(81)   a. encara que fos seguint algú més gran i *prenent nota de tot* el que feia.

(CTILC, 2004)

'even if it was by following someone older and taking notes of all that was done'

b. el gran bisbe Cesari d'Arle *va prendre la iniciativa d'escriure* una regla especial per a la comunitat.   (CTILC, 2004)

'the great bishop Cesari d'Arle took the initiative of writing a special rule for the community'

For Spanish *tomar*-LVCs, Figure 26 shows that low-frequency LVCs have the highest percentage of modified instances (35.25%), followed by high-frequency LVCs (30.78%); medium-frequency are the least modified (22.87%). These differences are statistically significant ($\chi2(2)=26.77$, $p<.0001$, Cramer's V=0.107, small effect).
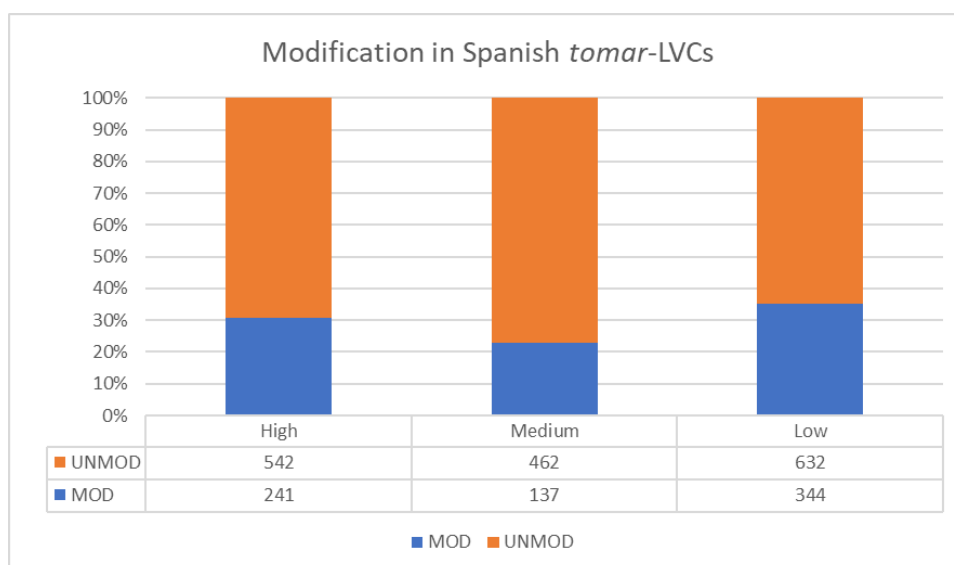
Figure 26. Distribution of modification in Spanish tomar-*LVCs according to the degree of frequency (high, medium or low)*

Table 33 shows that adjectival modification and PP modification display very similar frequencies in the three groups. The former is only more prevalent in medium-frequency *tomar*-LVCs: 12.35% of adjectival modification versus 9.35% of PP modification. In the case of high-frequency LVCs the proportions are very close: 13.54% of PP modification and 12% of adjectival modification. Likewise, in low-frequency LVCs there is a bit more modification, but both subclasses are very close: 16.60% of PP modification and 15.16% of adjectival modification.

Modification	High	Medium	Low
Unmod (No)	542 (69.22%)	462 (77.13%)	632 (64.75%)
Adj	94 (12%)	74 (12.35%)	148 (15.16%)
PP	106 (13.54%)	56 (9.35%)	162 (16.60%)
Rel	18 (2.30%)	5 (0.84%)	23 (2.36%)
NC	-	-	-
Gen	-	-	-
Various	23 (2.94%)	2 (0.33%)	11 (1.13%)
TOTAL	783	599	976

Table 33. Distribution of modification subclasses in *tomar*-LVCs according to the degree of frequency

The low percentages of adjectival modification in the results of Romance languages could be related to the high presence of bare nominals in the corpus examples

of this LV (as seen previously in §3.4.2.1). A closer analysis of each LVC will have its focus on the restrictions bare nominals have in Romance languages.

Like *give*-LVCs (§3.4), 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 and medium-frequency LVCs and a higher presence of modification in Germanic languages (Figure 23 and 24), but a negative correlation in Romance languages, where low-frequency LVCs present higher percentages of modification (Figure 25 and 26). Also, adjectival modification is more relevant in Germanic languages than in Romance languages, which favour PP modification with this verb. The rest of subclasses are, residual, as with the other verbs analysed.

The differences in these tendencies, as well as the effect of bare nominals in the presence or lack of modification in Romance languages will be seen in the upcoming sections with an analysis of each LVC selected for this sample.

3.5.2.2. English take-LVCs

As seen in Figure 23, high and medium-frequent English *take*-LVCs present a clear tendency to the modification of the nominal element, while less frequent LVCs have lower proportions of modification.

Figure 27 shows the distribution of modification in every particular *take*-LVCs analysed. The results prove that medium-frequency LVCs present the highest percentages of modification: *take a shower* (86.11%), *take a bite* (84.10%), *take a bath* (83.81%) and *take a nap* (77.50%), while *take a walk* (59.22%) and *take a risk* (40.91%) are modified below the mean which is 64.74% (see Figure 22). A similar behaviour can be found in high-frequency LVCs, where most LVCs are modified in the majority of instances: *take advantage* (85.28%), *take a breath* (71%), *take a picture* (63%), *take a step* (61.05%), and *take action* (60.80%). In contrast, low-frequency LVCs are the least modified: *take a drink* (52.92%), *take a stroll* (48.18%), *take a hike* (30%), and *take a swim* (21.43%).

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study
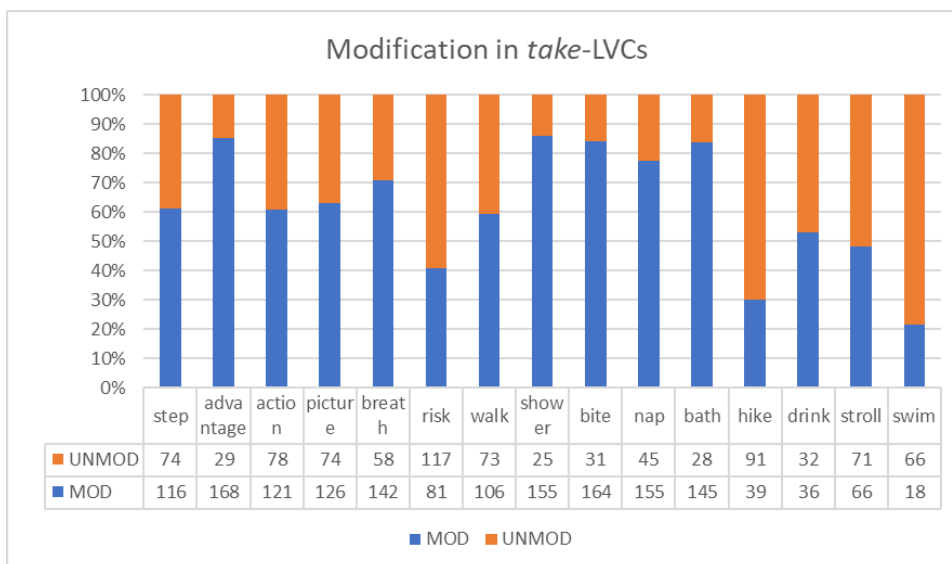
Figure 27. Distribution of modification in English take-*LVCs per lemma*

Table 34 shows the distribution of modification subclasses with a predominance of adjectival modification in all lemmas.

Modification	step	advantage	action	picture	breath	risk	walk	shower	bite	nap	bath	hike	drink	stroll	swim
Unmod (No)	74 (38.95%)	29 (14.72%)	78 (39.20%)	74 (37%)	58 (29%)	117 (59.09%)	73 (40.78%)	25 (13.89%)	31 (15.90%)	45 (22.50%)	28 (16.19%)	91 (70%)	32 (47.06%)	71 (51.82%)	66 (78.57%)
Adj	109 (57.37%)	18 (9.14%)	112 (56.29%)	73 (36.50%)	137 (68.50%)	63 (31.82%)	93 (51.96%)	150 (83.33%)	138 (70.77%)	115 (57,50%)	90 (52.02%)	28 (21.54%)	25 (36.77%)	48 (35.04%)	12 (14.29%)
PP	-	58 (29.44%)	-	15 (7.50%)	2 (1%)	3 (1.52%)	-	-	9 (4.62%)	-	-	-	3 (4.41%)	-	-
Rel	2 (1.05%)	5 (2.54%)	5 (2.51%)	1 (0.5%)	-	14 (7.07%)	-	-	-	-	-	-	1 (1.47%)	1 (0.73%)	-
NC	1 (0.53%)	1 (0.51%)	2 (1%)	14 (7%)	1 (0.5%)	-	13 (7.26%)	5 (2.78%)	-	38 (19%)	54 (31.21%)	10 (7.69%)	4 (5.88%)	17 (12.41%)	6 (7.14%)
Gen	-	-	-	4 (2%)	-	-	-	-	-	2 (1%)	-	1 (0.77%)	-	-	-
Various	4 (2.10%)	86 (43.65%)	2 (1%)	19 (9.5%)	2 (1%)	1 (0.5%)	-	-	17 (8.71%)	-	1 (0.58%)	-	3 (4.41%)	-	-
TOTAL	190	197	199	200	200	198	179	180	195	200	173	130	68	137	84

Table 34. Distribution of modification subclasses in *take*-LVCs per lemma

There is a clear exception for *take advantage* (9.14%) with a presence of adjectives that is lower than with PP modification (29.44%), as in (82). Apart from this, there are other LVCs with some presence of PP modification: *take a picture* (7.50%), *take a bite* (4.62%) and *take a drink* (4.41%), as in (83).

(82)   Before scientists can take *advantage of the assistance*, however, they need to critically assess both their personal goals   (COCA, 2012)

(83)   a. Before I go, can I just *take a few pictures of the girls* and maybe interview them?   (COCA, 2017)

b. Now take *one more bite of dessert* and get out of there.   (COCA, 2006)

c. Another helpful trick I learned is to take a take *a drink of water*.

(COCA, 2012)

Also, there is a relevant presence of noun classifiers with certain lemmas, especially *take a bath* (31.21%), *take a nap* (19%) and *take a stroll* (12.41%), but also *take picture* (7%), *take a walk* (7.26%), *take a hike* (7.69%), *take a drink* (5.88%), and *take a swim* (7.14%), as illustrated in (84). Finally, only one LVC shows modification by relative clauses: *take a risk* (7.07%), as in (85).

(84)   a. you had to take *a bubble bath*.   (COCA, 2017)

b. Cruz, who admittedly likes to take *an hour nap* prior to games,

(COCA, 2018)

c. In so much of my life, I take *sunset strolls* on the beach.

(COCA, 2012)

d. When did you take *that profile picture*?   (COCA, 2015)

e. The president himself struggled with insomnia, and was known to take *late night walks*.   (COCA, 2018)

f. who would want to take *an after-dinner drink* onto the deck and watch the sunset.   (COCA, 2000)

g. For a taste of civilization, take *a 2-mile hike* over to Bascom Lodge for dinner.   (COCA, 2004)

h. She would take *her morning swims* and read a book underneath the shade of a palm.   (COCA, 2014)

(85)   then we don't prepare them to take *the kind of risks they need to take* to be disciples and to have joyful and fulfilling lives.   (COCA, 2010)

In the cases of the most modified English *take*-LVCs through adjectival modification, an examination of the co-occurrence with adjectives shows a tendency to collocate with specific adjectives. For instance, *take a shower* shows a low variety in the combination because it co-occurs with four adjectives in the majority of modified cases (91/150, 60.67%), as in (86), which are related to kinds of showers, that is, *hot* (38/150, 25.33%), *cold* (18/150, 12%), *quick* (23/150, 15.33%) and *long* (12/150, 8%).

(86)   a. Never take *a hot shower* afterwards.   (COCA, 2016)

b. And by all means, take *a cold shower*. It may make you cleaner, but it won't sober you up   (COCA, 2012)

c. It was just the regular day. I went to take *a quick shower*.

(COCA, 2017)

d. - Take *an extra-long shower*? - Go upstairs. (COCA, 2003)

A similar pattern is found in *take a bite* with a majority of modified instances co-occurring with four adjectives (85/155, 54.84%): *big* (30/155, 19.55%), *first* (23/155, 14.84%), *small* (18/155, 11.61%) and *huge* (14/155, 9.03%), as in (87).

(87)　　a. Wouldn't you just love to take *a big bite* of that, Elinor? (COCA, 2012)

b. Those guys must sit down to dinner and can't even take *the first bite* without laughing their asses off at how dumb people are. (COCA, 2012)

c. Then take *a small bite* of the chip and notice how it tastes.

(COCA, 2008)

d. sees one of them grab the candy bar from the boy and take *a huge bite* out of it. (COCA, 2009)

To a lesser extent, *take a nap* also presents a co-occurrence with mainly four adjectives in more than half of the instances modified by an adjective (72/115, 62.61%), as in (88): *little* (38/115, 33.04%), *long* (16/115, 13.91%), *quick* (9/115, 7.83%) and *short* (7/115, 6.09%).

(88)　　a. I just want to take *a little nap*. Just a really short nap. (COCA, 2013)

b. go outside and stare into the sun for a while and then take *an extra-long nap*. (COCA, 2012)

c. I took advantage of the time to myself to take *a quick nap*.

(COCA, 2012)

d. maybe shed bring him home, feed him, he could take *a short nap*.

(COCA, 2010)

Moreover, there are two cases of *take*-LVCs which present a high co-occurrence with a single adjective. First, *take a breath* is mostly modified by the adjective *deep* (96/139, 69.06%), as in (89). Second, *take a walk* co-occurs in many instances with the adjective *long* (46/93, 49.46%), as in (90).

(89)　　Here, just take *deep breaths*. Okay. Deep breaths, all right?

(COCA, 2014)

(90)   I always take *a really long walk* when I feel damp and drizzly in my soul..

(COCA, 2012)

In contrast, *take a bath,* does not have a clear tendency to co-occur with specific adjectives, and the co-occurrence with similar adjectives as *take a shower* is lower: *hot* (14/91, 15.38%), *long* (13/91, 14.29%) and *warm* (9/91, 9.89%), as in (91).

(91)   a. Do not take *a hot bath* or a detox bath if you are pregnant

(COCA, 2012)

b. Then they can just go away and take *a super long bath*

(COCA, 2012)

c. Why don't you take *a warm bath*? It'll help with the cramps.

(COCA, 2018)

With regards to *take action*, the variety of adjectives is wide and there are no patterns of high co-occurrence with certain adjectives (92), unlike the rest of high-frequency *take*-LVCs.

(92)   a. But they've agreed not to take *any further legal action* against you or
against Andrews Prep   (COCA, 2008)
b. the realization that we need to take *clear and decisive action* to ensure
the viability of our Catholic schools.   (COCA, 2008)
c. Some of us provide information... others give their support. Some take
more direct action. Like you.   (COCA, 2002)

In a similar line, the less modified LVCs do not show tendencies to co-occur with specific adjectives. On the contrary, when there are few examples, the adjectives are barely repeated and the range of qualities they refer to is wide, as in *take a hike* (93), *take a drink* (94), or *take a swim* (95).

(93)   a. Take *a leisurely hike* in the woods, find a couple of mushrooms

(COCA, 2013)

b. New York Scouts take *a historic hike* along a channel of water more than
120 MILES LONG   (COCA, 2012)

c. I thought maybe tomorrow we could take *a romantic hike*.

(COCA, 2012)

(94)   a. Come on, come on. One drink. Take *a little drink.*   (COCA, 2012)

b. as if he were leaning into a fountain to take *a long drink.* (COCA, 2015)

(95)   a. I'm just going to take *a little swim.*   (COCA, 2019)

b. If you want to take *a long swim* underwater, the trick is to breathe in and out a few times   (COCA, 2013)

c. They are on the beach, and the man is anxious to take *one last swim* before they have to leave   (COCA, 2012)

The preference of these LVCs for adjectival modification also corresponds to the general tendency of the language nominal domain, and only PP modification and noun classifiers are relevant for *take*-LVCs.

3.5.2.3. German nehmen-LVCs

Figure 28 presents the results of adjectival modification for each German *nehmen*-LVC, which is the least modified verb (5.71%, see Figure 22 above). The most modified LVCs are *Einfluss nehmen* (16.56%), a high-frequency LVC, and *Anteil nehmen* (13.91%), a medium-frequency LVC. The rest of lemmas present much lower percentages: *Rücksicht nehmen* (8.38%), *Einsicht nehmen* (2.22%), *Anstoß nehmen* (1.02%) and *Abschied nehmen* (0.66%). In fact, there are two LVCs which are not modified in any of the extracted instances: *Platz nehmen* and *Abstand nehmen.*

	Abschied	Rücksicht	Einfluss	Platz	Abstand	Einsicht	Anteil	Anstoß
UNMOD	151	153	131	175	175	44	130	97
MOD	1	14	26	0	0	1	21	1
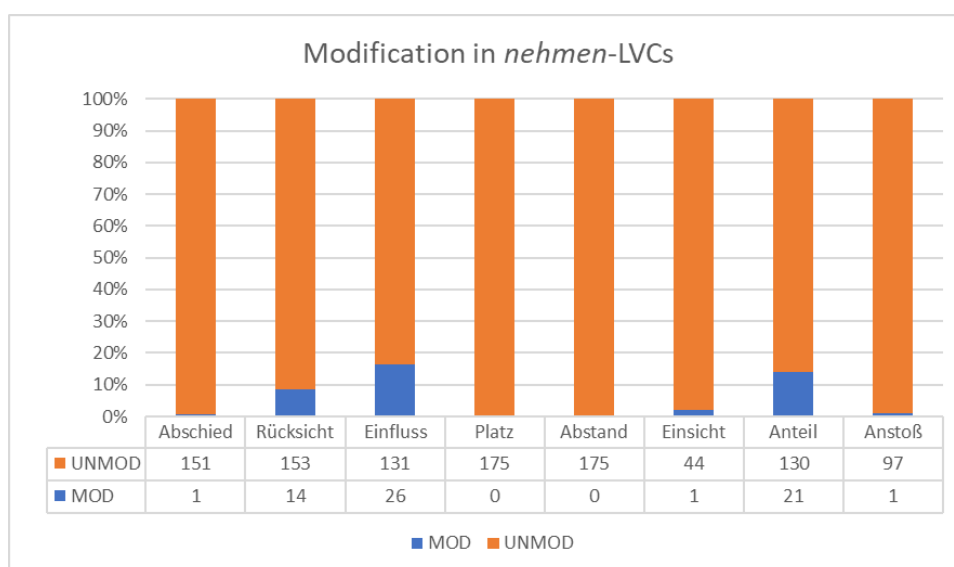
Figure 28. Distribution of modification in German nehmen-*LVCs per lemma*

When the co-occurrence with adjectives is examined, the three LVCs which present higher percentages of adjectival modification do not show any tendency to co-

occur with specific adjectives. Right the opposite, the variety of adjectives which co-occur with these LVCs is wide and it includes both qualitative (96) and relational adjectives (97).

(96) a. Leider hat hier das amerikanische Rechtssystem *einen unheilvollen Einfluss* genommen.   (DWDS, 2003)

‘Unfortunately the American law system has taken an ominous influence here’

b. Die neuen Heiligen haben *lebendigen Anteil* genommen am Leben der Menschen   (DWDS, 2014)

‘The new saints have taken a vivid interest in people’s life’

c. "Auf Familien wird durch Freibeträge *besondere Rücksicht* genommen
                     (DWDS, 2003)

‘Tax allowances is taken into great consideration by families’

(97) a. da können wir mit Sicherheit auf einen Demokratisierungsprozess beispielsweise in Ägypten, in Tunesien oder in anderen Ländern *einen sehr positiven Einfluss* nehmen.   (DWDS, 2012)

‘since we could take with security a very positive influence in a process of democratization in Egypt, Tunisia and other countries’

b. Gumilev nahm an diesem Spektakel *aktiven Anteil...* (DWDS, 2002)

‘Gumilev took active part in this spectacle’

c. Und Peter Daniel fürchtet, die Stiftung Preußische Schlösser und Gärten müsse *politische Rücksichten* nehmen.   (DWDS, 2001)

‘And Peter Daniel feared that the Institute could take Prussian castles and gardens into political consideration’

In the case of *nehmen*-LVCs, however, there also LVCs which are not modified by adjectives at all, or which are in very low percentages, as in (98) where there is only one instance with adjectival modification.

(98) a. Vielleicht hat er auch schon an diesem Tag *seinen persönlichen Abschied* genommen.   (DWDS, 2012)

‘Maybe he also resigned personally in the same day’

b. Das Zweite: Sie wollen, dass der Rechnungshof in Zukunft *bessere Einsicht* nehmen kann.   (DWDS, 2019)

'The second: They want that the Auditor's Department will make a better inspection in the future'

c. *Besonderen Anstoß* nahmen sie   (DWDS, 2015)

'They took relevant offence'

These low results are linked to a vast majority of bare singulars with *nehmen*-LVCs (Figure 16). In some cases, LVCs with *nehmen* seem to conform a unit which occupy a single position in the sentence, as in (99) where the whole construction (*Anstoß nehmen*) is in the first position followed by the modal verb *muss* which is in the second position, as German is a strict V2 language. However, German is also a language that can front the whole VP, or even just a part of the VP in partial fronting (Müller, 2022), so this is not particular to LVCs but to the language in general.

(99)   *Anstoß nehmen* muss man dennoch an des Trainers Fingerübungen -

(DWDS, 2000)

'One must still take offence on the finger exercises of the coach'

In other cases, the unit they conform is made clear by the adjacency of the NVE and the LV in embedded clauses in German (100), where the verb should occupy the last position in the sentence.

(100)   a. Ich erinnere mich an "Die Erfindung der Einsamkeit", das Buch, in dem Paul Auster *Abschied nimmt* von seinem Vater   (DWDS, 2012)

'I remember "The Invention of Solitude", the book in which Paul Auster takes leave from his father'

b. Erst wenn wir *Einsicht nehmen* in die Grundstruktur unseres Scheiterns,

(DWDS, 2002)

'Only when we make inspection of the our failures' basic structure'

c. Die Akten zeigen Zimmermann zufolge, "dass die drei Suchende waren, Menschen, die zweifelten und auch *Anstoß nahmen* an den Verhältnissen – so wie viele junge Menschen".   (DWDS, 2016)

'The act shows Zimmermann the following: "that the three searchers were people that doubted and also took offence on the relationships, like many young people"'

In sum, the verb *nehmen* selects a vast majority of bare singulars with no modification, of any kind. These results emphasize its special behaviour in comparison

144

to the previous verb, *geben*-LVCs, as well as the upcoming verb, *machen*-LVCs, discussed in section 3.6.2.3.

3.5.2.4. Catalan prendre-LVCs

Catalan *prendre*-LVCs have shown a tendency to be modified in around a third of instances in all groups according to the degree of frequency (see Figure 25).

Taking into account that the average number of modified *prendre*-LVCs in Catalan is 30.02% (see Figure 22), Figure 29 shows that there is an irregular distribution of adjectival modification in the different lemmas. There are some LVCs with higher percentages of modification in relation to a low number of instances, such as *prendre molèstia* (78.38%), *prendre possessió* (65.71%), *prendre una resolució* (54.55%), *prendre nota* (39.57%), *prendre una decisió* (50%), *prendre determinació* (38.89%), *prendre impuls* (33.33%), and *prendre bany* (33.33%). The rest of LVCs present lower percentages than the average: *prendre precaució* (11.54%), *prendre distància* (7.41%), *prendre iniciativa* (21.95%), *prendre consciència* (4.76%) and *prendre partit* (4.55%). There is also a particular LVC which is not modified in any of the extracted instances: *prendre represàlia*.

	decisió	consciència	nota	possessió	iniciativa	partit	precaució	molèstia	distància	determinació	bany	resolució	impuls	represàlia
UNMOD	21	100	58	12	32	42	23	8	25	11	14	5	4	4
MOD	21	5	38	23	9	2	3	29	2	7	7	6	2	0
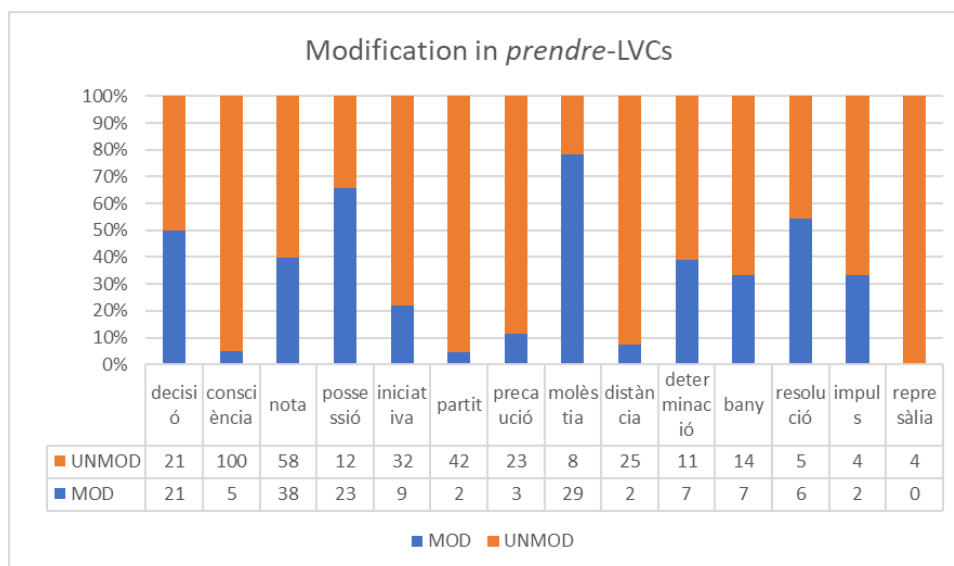
Figure 29. Distribution of modification in Catalan donar-*LVCs per lemma*

Table 35 shows that the distribution of subclasses of modification in Catalan *prendre*-LVCs is not varied: most LVCs favour adjectival modification, although this is not high, and some NVE favour PP modification over adjectives: *prendre nota* (30.21% of PP modification versus 3.12% of adjectival), *prendre molèstia* (72.97% of PP modification versus 5.41% of adjectival), *prendre bany* (23.81% of PP modification

versus 9.52% of adjectival), and even *prendre possessió* only presents modified instances through PP modification (65.71%).

Modification	decisió	consciència	nota	possessió	iniciativa	partit	precaució	molèstia	distància	determinació	bany	resolució	impuls	represàlia
Unmod (No)	21 (50%)	100 (95.24%)	58 (60.43%)	12 (34.29%)	32 (78.05%)	42 (95.45%)	23 (88.46%)	8 (21.62%)	25 (92.59%)	11 (61.11%)	14 (66.67%)	5 (45.45%)	4 (66.67%)	4 (100%)
Adj	11 (26.19%)	5 (4.76%)	3 (3.12%)	-	3 (7.31%)	2 (4.55%)	2 (7.69%)	2 (5.41%)	2 (7.41%)	4 (22.22%)	2 (9.52%)	5 (45.45%)	2 (33.33%)	-
PP	10 (23.81%)	-	29 (30.21%)	23 (65.71%)	4 (9.76%)	-	1 (3.85%)	27 (72.97%)	-	2 (11.11%)	5 (23.81%)	1 (9.10%)	-	-
Rel	-	-	3 (3.12%)	-	2 (4.88%)	-	-	-	-	1 (5.56%)	-	-	-	-
NC	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Gen	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Various	-	-	3 (3.12%)	-	-	-	-	-	-	-	-	-	-	-
TOTAL	42	105	96	35	41	42	26	37	27	18	21	11	6	4

Table 35. Distribution of modification subclasses in *prendre*-LVCs per lemma

It is important to note that the PP found in these constructions are of different nature, as illustrated in (101). Following Cremades' analysis (2014, 2017) for the argument selection in Catalan noun phrases, we can find instances of different PP modifiers. Some Romance PP modifiers can be assimilated to the genitive in Germanic languages (as pointed out in Bonet & Solà, 1986), as in *un bany de mar* (101c). Others can be considered quasi-arguments holding different thematic roles (as proposed in Levin & Rappaport Hovav, 2005), such as Patient, Theme or Goal, as in *possessió dels seus cors* (101d). Further, some instances are attributive PPs introducing qualities and a description of the noun, as in *determinacions d'aquesta mena* (101e).

(101)   a. Va prendre *nota de la documentació personal de les dues*.

(CTILC, 2012)

'S/he took note of the personal documentation of both'

b. ¿qui es prendrà *la molèstia de refutar-ho*?   (CTILC, 2000)

'who will take the time to reject it?'

c. i tant fa si l'oratge, bo o dolent, permet prendre *un bany de mar*.

(CTILC, 2012)

'and it does not matter if the weather, good or bad, allows to take a sea dip'

d. tot esperant que la pau divina hagués pres *possessió dels seus cors* i de les seves coses.   (CTILC, 2004)

'waiting that the holy peace would have taken possession of their hearts and their things'

e. però molt intel·ligent per prendre *determinacions d'aquesta mena*?

(CTILC, 2005)

'but very intelligent to take determinations of this kind?'

The LVC with the highest number of modified instances by adjectives (*decisió*) shows a wide variety of adjectives, both relational (102a) and qualitative (102b,c,d). This flexibility in accepting adjectival modification could be related to the determiner requirement imposed by this LVC, which contrasts with the tendency to co-occur with bare nouns displayed by the rest of *prendre*-LVCs.

(102)   a. que pren *les decisions polítiques*;   (CTILC, 2010)
'that makes the political decisions'
b. qui pren *les decisions transcendents*   (CTILC, 2013)
'who takes the important decisions'
c. donant-los tota la informació que necessiten per prendre *les decisions més encertades*   (CTILC, 2011)
'giving them all the information that they need to make the most accurate decisions'
d. En casos de vacil·lació com aquests cal prendre *decisions coherents i sistemàtiques* en fer la transcripció dels registres   (CTILC, 2011)
'In hesitating cases like these one must make coherent and systematic decisions when transcribing the registers'

In other cases, there are certain restrictions to adjectival modification which are related to the bareness of the nominal element. In the case of *prendre impuls* (103) the

two instances with adjectival modification are combined with the adjective *nou* ('new),
while the case of *prendre bany* (104) combines with the adjectives *calent* ('hot') and *llarg*
('long'), which are the adjectives already detected in the case of English *take a bath*.

(103)	la vivacitat de la conversa prenia *un nou impuls*.	(CTILC, 2001)
	'the vigour of the conversation took a new push'	
(104)	a. m'ha dit que havia de prendre *un bany calent*.	(CTILC, 2015)
	'she told me that I had to take a hot bath'	
	b. Vaig prendre *un bany bastant llarg*	(CTILC, 2017)
	'I took a quite long bath'	

The low number of instances for *prendre*-LVCs, and the even more limited
instances with adjectival modification, limits the generalisations of this verb for this
language. However, the behaviour of PP modification has a greater relevance for this
verb, and it contrasts with the modification in the other two LVs in Catalan: *donar* (§3.4)
and *fer* (§3.6).

3.5.2.5. Spanish tomar-LVCs

Figure 30 shows that the irregular tendency detected in the previous section on
Catalan is further confirmed in the analysis of each LVC in the other Romance language
under study. While the mean of modification cases for *tomar*-LVCs is 30.62% (see Figure
22), there are a few lemmas that are above this percentage from among the high frequency
group (*tomar una decisión,* 49.22%; and *tomar nota,* 39.50%), the medium-frequency
group (*tomar precaución*, 38.19%) and especially the low frequency group (*tomar
determinación*, 57.87%, *tomar baño,* 45.73%; and *tomar atajo,* 31.91%). The rest of
LVCs have lower percentages of adjectival modification: *tomar descanso* (28.28%),
tomar acuerdo (27.68), *tomar fotografía* (25.26%), *tomar iniciativa* (23.50%), or even
much lower, as in *tomar represalia* (12.50%), *tomar conciencia* (9.5%), and *tomar
distancia* (7%).
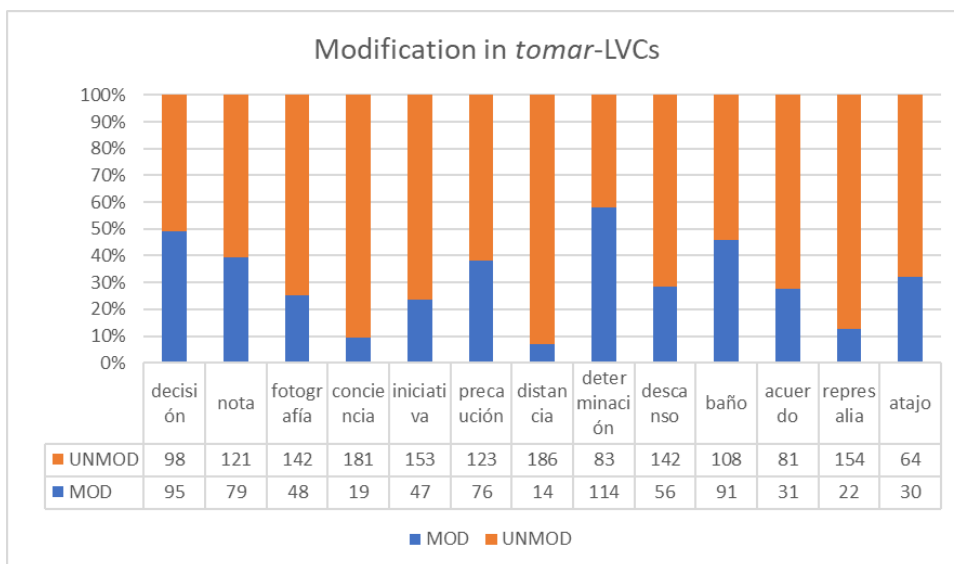
Figure 30. Distribution of modification in Spanish tomar-*LVCs per lemma*

Table 36 shows that the distribution of subclasses of modification in Spanish *tomar*-LVCs prefers adjectival modification in 8 lemmas, while the other 5 favour PP modification.

Modification	decisión	nota	fotografía	conciencia	iniciativa	precaución	distancia	determinación	descanso	baño	acuerdo	represalia	atajo
Unmod (No)	98 (50.78%)	121 (60.50%)	142 (74.74%)	181 (90.50%)	153 (76.50%)	123 (61.81%)	186 (93%)	83 (42.13%)	142 (71.72%)	108 (54.27%)	81 (72.32%)	154 (87.50%)	64 (68.09%)
Adj	52 (26.94%)	9 (4.50%)	22 (11.58%)	11 (5.50%)	14 (7%)	49 (24.62%)	12 (6%)	26 (13.20%)	41 (20.71%)	41 (20.60%)	10 (8.93%)	20 (11.36%)	10 (10.64%)
PP	26 (13.47%)	57 (28.50%)	20 (10.53%)	3 (1.50%)	31 (15.50%)	24 (12.06%)	1 (0.5%)	75 (38.07%)	15 (7.57%)	45 (22.61%)	19 (16.96%)	1 (0.57%)	7 (7.44%)
Rel	9 (4.66%)	2 (1%)	2 (1.05%)	5 (2.50%)	2 (1%)	2 (1.01%)	1 (0.5%)	7 (3.55%)	-	-	2 (1.79%)	1 (0.57%)	13 (13.83%)
NC	-	-	-	-	-	-	-	-	-	-	-	-	-
Gen	-	-	-	-	-	-	-	-	-	-	-	-	-
Various	8 (4.15%)	11 (5.50%)	4 (2.10%)	-	-	1 (0.5%)	-	6 (3.05%)	-	5 (2.52%)	-	-	-
TOTAL	193	200	190	200	200	199	200	197	198	199	112	176	94

Table 36. Distribution of modification subclasses in *tomar*-LVCs per lemma

The *tomar*-LVCs preferring PP modification show a clear distribution of this class of modification in comparison to adjectival modification: *tomar nota* (28.50% of PP modification versus 4.50% of adjectival), *tomar iniciativa* (15.50% of PP modification versus 7% of adjectival), *tomar determinación* (38.07% of PP modification versus 13.20% of adjectival), *tomar baño* (22.61% of PP modification versus 20.60% of adjectival), and *tomar acuerdo* (16.96% of PP modification versus 8.93% of adjectival), as illustrated in (105).

(105)   a. un vigilante, planilla y reloj en mano, tomaba *nota de su paso*.

                                    (CORPES XXI, 2002)

'a guard, with a spreadsheet and a watch in his hand, took note of his passing by'

b. Asegúrales que nosotros tomaremos *la iniciativa de una nueva entrevista*.   (CORPES XXI, 2002)

'Make sure than we will take the initiative in a new interview'

c. por lo tanto, con base a la experiencia y el saber de los productores se debe tomar *la determinación de la densidad a emplear*.

                                      (CORPES XXI, 2003)

'therefore, with the basis of the experience and knowledge of producers one must take the determination of the density to be used'

d. Si Penélope y Cleopatra, según la leyenda, tomaban *baños de leche* para embellecerse   (CORPES XXI, 2004)

'If Penelope and Cleopatra, according to the legend, took milk baths to embellish'

e. para tomar *acuerdo sobre algún punto contenido en agenda* los estatutos exigen mayoría calificada.   (CORPES XXI, 2004)

'to come to an agreement about any point in the agenda, the statutes require a qualified majority'

Also, some of the other lemmas also present a considerable proportion of PP modification, even if the adjectival modification is predominant: *tomar decisión* (13.47%), *tomar fotografía* (10.53%), *tomar precaución* (12.06%), *tomar descanso* (7.57%), and *tomar atajo* (7.44%), as in (106). Finally, there is also modification through

relative clauses in most LVCs with special relevance in *tomar decisión* (4.66%), and *tomar atajo* (13.83%), as in (107).

(106)   a. Michael tomó *la decisión de seguir adelante* no sin mantenerse alerta.

(CORPES XXI, 2001)

'Michael made the decision of continuing, not without being alert'

b. Lianeth tomaba *fotografías de todas partes*. (CORPES XXI, 2003)

'Lianeth took pictures of everywhere'

c. y era necesario tomar *las precauciones del caso*.

(CORPES XXI, 2003)

'and it was necessary to take the precautions of the case'

d. tomará *un descanso de varias semanas*,   (CORPES XXI, 2007)

'he will take a several-week rest'

e. por lo que decidieron tomar *un atajo de 2 kilómetros* saliéndose de la carretera principal.   (CORPES XXI, 2015)

'that's why they decided to take a 2 kilometres shortcut outside of the main road'

(107)   a. así tomar *las decisiones que para el caso a ellos les parezcan pertinentes*.   (CORPES XXI, 2001)

'so taking the decisions that they found appropriate for the case'

b. deciden tomar *un atajo que ven en su teléfono celular*.

(CORPES XXI, 2005)

'they decide to take a shortcut that they find in their cellphone'

As in the case of Catalan *prendre*, the PP modification has a special relevance in *tomar*-LVCs, and PP modifiers of different nature have been found: genitive PPs, as in *baños de leche* (105d); quasi-argumental PPs (as defined by Cremades, 2017, for Catalan), as in *fotografías de todas partes* (106b); as well as attributive PPs, as in *un atajo de 2 kilómetros* (106e).

When the co-occurrence with adjectives is examined, a tendency to co-occur with specific adjectives has been found in three lemmas. This is the case of *tomar precaución* ('take caution') with *necesaria* ('necessary') in 17/50 instances (34%) and *debida* ('due') in 7/50 instances (14%), as in (108); as well as *tomar baño* ('take bath') with *caliente* ('hot') in 10/46 instances (21.74%), as in (109), and *tomar descanso* ('take rest') with

merecido ('deserved') in 9/41 instances (21.95%), *pequeño* ('little') in same number of instances (21.95%) and *breve* ('brief') in 8/41 instances (19.55%) as in (110).

(108)   a. No se podría acusar a nadie: se habían tomado *las precauciones necesarias* para proteger a la aldea.   (CORPES XXI, 2002)

'Nobody could be blamed: all due precautions were taken in order to protect the village'

b. Es un país seguro para el viajero, lo cual no quiere decir que no deban tomarse *las debidas precauciones* para prevenir el delito

(CORPES XXI, 2002)

'It is a safe country for the traveller, which does not mean that due precautions should not be taken to prevent the criminal offense'

(109)   No, no ahora, primero tomaré *un baño bien caliente*.

(CORPES XXI, 2001)

'No, not now, I will first take a very hot bath'

(110)   a. me tomaré *un merecido descanso* en un rincón. (CORPES XXI, 2001)

'I will take a deserved rest in a corner'

b. y nos tomáramos *un pequeño descanso* antes de volver a comenzar.

(CORPES XXI, 2002)

'and we would take a tiny break before restarting'

c. Tomó *un breve descanso* antes de viajar a Guayaquil para el concierto que brindará esta noche.   (CORPES XXI, 2003)

'He took a short break before travelling to Guayaquil for the concert that he will give tonight'

The rest of LVCs with higher percentages of adjectival modification do not show any tendency to co-occur with specific adjectives. This includes a great variety of qualitative adjectives (111), but also some relational adjectives (112).

(111)   a. Probablemente percibirán que las personas que los rodean toman *decisiones inexplicables e incluso contradictorias.* (CORPES XXI, 2001)

'You will probably understand that people who surround them make inexplicable and even contradictory decisions'

b. no toma *determinaciones drásticas* porque espera que la Iglesia y la Alcaldía medien en el tema.   (CORPES XXI, 2005)

'he does not take drastic determinations because he hopes that the Church and the Major will intercede in the matter'

c. Y fiel a sus peores fantasías, tomó *una determinación más torpe* que todas sus otras, y famosas, determinaciones torpes.

(CORPES XXI, 2002)

'And faithful to his/her worst fantasies, he/she took a clumsier determination than his/her other well-known clumsy determinations'

(112)   a. las reglas sociales que permiten tomar *decisiones alimentarias*.

(CORPES XXI, 2001)

'social rules that allow to make food decisions'

b. Esas eran las ideas de los que tomaron *la decisión política*,

(CORPES XXI, 2001)

'Those were the ideas of those who make the political decision'

c. El general Ricardo Izurieta y un puñado de generales tomaron *la decisión histórica* de reconocer abiertamente ante el país lo ocurrido.

(CORPES XXI, 2001)

'General Ricardo Izurieta and a handful of generals made the historical decision of openly acknowledging what happened in front of the country'

The low percentage of modification of this LV is in line with the previous languages, Catalan and German, and contrasts with the results obtained in the examination of the English language. Unlike the case of Catalan *prendre*-LVCs, the optionality of the determiner in *tomar*-LVCs does not present restrictions in their selection of the types of adjectives.

3.5.3. Discussion

The LV *take* shows clear differences with the other two LVs *give* and *make* in all languages. The divergences in determination, for instance, have proven to be statistically significant, with English *take*-LVCs presenting the highest percentage of modified NVEs (74.43%) followed by Spanish *tomar*-LVCs with just above half of instances with a determiner (56.78%), contrasting with Catalan *prendre*-LVCs with a third of instances (34.70%) and especially with German *nehmen*-LVCs with a very low proportion of determined NVEs (12.86%). The main pattern observed in this verb is that it tends to select more bare nominals, except in English. In English *take*-LVCs, almost all LVCs select determined NVEs with two exceptions: *take advantage* and *take action*, which

153

combine with bare singulars. These results, which stand out from the general tendency of the rest of the lemmas, could be taken as signs of lexicalization in these two LVCs. In contrast, German *nehmen*-LVCs combine with bare singulars in all lemmas, which contrasts with the other Germanic language (English), but also with the other two German LVs analysed (*geben* and *machen*). In the case of the two Romance languages, a similar trend is found: bare singulars are mostly found in high-frequency LVCs, while low-frequency LVCs present higher percentages of determination. Some of the most common lemmas with bare singulars are shared between both languages (i.e. *prendre consciència / tomar conciencia* 'get awareness', and *prendre nota / tomar nota* 'take note'). Once again, some Spanish LVCs combine with bare plurals more generally (i.e. *tomar represalias* 'to retaliate') than in Catalan.

When the correlation between determination and modification is examined in general terms, more modification in English bare nouns is found than in the rest of the languages, although English *take*-LVCs has the lowest proportion of bare nominals by a significant difference. In the case of Catalan *prendre*-LVCs and Spanish *tomar*-LVCs, some signs of lexicalization in the high-frequency lemmas analysed can be traced, which have been also contrasted with the different subclasses of modification. However, the case of German *nehmen*-LVCs shows significant differences with the other two verbs analysed because its tendency to combine with bare nominals in object position does not follow the general patterns of the language (Chierchia, 1998).

Once again, the results present lower instances of modification than the previous LV *give*. In the case of *take*-LVCs, there is a clear-cut distinction between the tendencies in English, with 64.74% of modification, and the rest of the languages: 30.62% in Spanish, 30.02% in Catalan and only 5.71% in German (see Figure 22).

These results strongly contrast with the previous research by Levin & Ström Herold (2015) for Germanic languages, which presented a lower presence of adjectival modification for the English *take*-LVCs (22%) and a higher percentage for German *nehmen*-LVCs (5.8%). These percentages, however, should be taken cautiously because their number of instances is much lower (up to 265 instances) than those of the present study (more than 1,000 examples in both Germanic languages). Interestingly, my results for German are in line with their results, but they differ in the case of English *take* which is modified in 47.86% of the instances by an adjective (Table 34).

Regarding Bonial's (2014) study of the LV *take,* modification is more frequent in the case of LVCs than lexical/synthetic verbs, and it is pointed out that it is qualitatively different. Their results of modification are also above 60% in instances of *take*-LVCs analysed in the PropBank corpus, which is a similar proportion to the one in my sample. In her study, however, Bonial includes other classes under the modification label such as quantifiers and determiners (which are studied under the determination label in this dissertation, as justified in §3.2.3).

In the case of Romance languages, the results of Spanish and Catalan present similar percentages to those found in the previous section with the verb *give*. For *take*, Spanish *tomar*-LVCs show 30.62% of modification, and Catalan *prendre*-LVCs 30.02%. These results cannot be contrasted to previous studies because the NVE have not been studied on the basis of corpora for these languages. What is relevant in the Romance data for this LV is that it selects PP modification in a larger proportion than adjectival modification in Catalan *prendre*-LVCs (8.38% and 19.88%, respectively); while the proportions are almost equal in Spanish *tomar*-LVCs (13.44% and 13.74%, respectively).

It is important to bear in mind that my results show very little prominence of modification subclasses other than adjectival modification in all four languages, and PP modification in the two Romance languages. This is also in line with previous research by Levin & Ström Herold (2015: 24) who discard the in-depth analysis of clausal modifiers because they represent only 1% of instances in *take*-LVCs and 0 instances in *nehmen*-LVCs. The proportions are consistent with my data: in English *take*-LVCs, relative clauses represent 1.15%; while there is no relative clause modification in German *nehmen*-LVCs. In the case of Romance languages, relative clauses also show residual presence: in Catalan *prendre*-LVCs, 1.17%; and in Spanish *tomar*-LVCs, 1.95%.

When the focus is placed on the degree of frequency of these LVCs in corpora, there are tendencies which set apart the Germanic and Romance languages under study. The former, English and German, show a stronger presence of modification in medium-frequency LVCs as well as with high-frequency lemmas, but it is significantly less present in low-frequency LVCs. This tendency is reversed in the Romance languages, Catalan and Spanish, where the most modified LVCs are those belonging to the low-frequency group. This tendency is made clear in the case of Catalan, while the Spanish percentages of medium and high-frequency LVCs are closer to low-frequency LVCs.

In the study of particular LVCs in all four languages, only English and Spanish present a clear co-occurrence of certain adjectives with the most modified LVCs (i.e., *take a deep breath* in 70.07% of modified instances, *take a little nap* in 33.04% or *take a hot shower* in 25.33%; *tomar necesaria precaución* 'take necessary precaution' in 34.60% or *tomar baño caliente* 'take a hot bath' in 24.39%), which could have influenced the overall percentage of adjectival modification. At the same time, the least modified LVCs do not show co-occurrence with any specific adjective. This contrasts with the case of German and Catalan where no actual co-occurrence of specific adjectives has been found, which could be explained by the low number of modified instances overall in both languages with very low percentages.

In general, the variety of adjectives is wide in most lemmas and both qualitative and relational adjectives are present. In fact, the restrictions on modification found in bare nominals, especially in Romance languages, are not present for this LV although the selection of bare nominals is high.

3.6. Quantitative results: the LV *make*

The light verb *make* is also one of the basic LVs from a crosslinguistic perspective (Butt, 2010). In present day English, the LV *make* is widespread, and its study alongside the LV *take* is interesting due to their coexistence in different varieties of the language (i.e. *make/take a decision*; Colominas, 2001). For German, the verb *machen* has been classified as the most common German LV (Bruker, 2011: 45). Also, *hacer* and *fer* are very present in the Spanish and Catalan LVCs, respectively (Sanromán Vilas, 2014; Ginebra & Navarro, 2015; De la Cruz, 2021). In fact, there is an important difference in the relevance of this verb in these two Romance languages, because Catalan LV *fer* is the default LV in the language, as can be deduced from its productivity and high frequency (Rosselló, 2002; GIEC, 2016).

Table 37 (repeated here for convenience) shows LVCs with the verb *make* in the four languages under study. The selection of LVCs is based on the frequency list in the different corpora, as well as on information from previous studies: for English, Perea Irigoyen & Sánchez Róo (2013) and Mehl (2018); for German, Bruker (2011: 45); for Spanish, Alonso-Ramos (2004: 325-26); for Catalan, *GIEC* (17.2.2) and De la Cruz (2021).

	High frequency	Medium frequency	Low frequency
Make (English)	decision, mistake, choice, call, progress	claim, contribution, profit, payment	adjustment, announcement, assumption, accusation, mention
Machen (German)	Gedanke ('thought'), Sorge ('worry'), Fehler ('mistake'), Angabe ('detail')	Erfahrung ('experience'), Hoffnung ('hope'), Anfang ('beginning')	Foto ('picture'), Sport (sport'), Ausführung ('implementation'), Beobachtung ('observation'), Eindruck ('impression')
Fer (Catalan)	Referència ('reference'), Esforç ('effort'), Feina ('work')	Ús ('use'), Pregunta ('answer'), Visita ('visit'), Viatge ('trip'), Petó ('kiss')	Anàlisi ('analysis'), Dibuix ('drawing'), Fotografia ('picture'), Abraçada ('hug'), Aclariment ('clarification')
Hacer (Spanish)	Pregunta ('answer'), Esfuerzo ('effort'), Referencia ('reference'), Uso ('use')	Viaje ('trip'), Ejercicio ('exercise'), Análisis ('analysis'), declaración ('statement')	Carrera ('run'), Énfasis ('emphasis'), Crítica ('review'), Broma ('joke'), Colección ('collection')

Table 37. *Make*-LVCs analysed in the four languages under study

The final database of *make*-LVCs instances consists of 8,019 tokens of LVC with some differences in the distribution among the four languages: Spanish and English provide the highest number of examples, above 2,000 each, while German and Catalan instances are about 1,000 fewer (Table 38). These divergences are consistent with the differences presented in the size of the corpora (seen in 3.2.2.).

	English	German	Catalan	Spanish	TOTAL LV
Make | Machen | Fer | Hacer	2,687	1,361	1,541	2,430	8,019

Table 38. Distribution of the number of LVCs of the database for each language

As in previous LVs, both determination (§3.6.1) and modification (§3.6.2) in *make*-LVCs are presented in the upcoming subsections. After the presentation of the results in details, a discussion will follow to contrast and compare the findings (§3.6.3).

3.6.1. Determination in *make*-LVCs

Figure 31 shows that determination is predominant in English *make*-LVCs (82.58%), followed by the rest of the languages with very close results: Catalan *fer*-LVCs (69.76%), German *machen*-LVCs (57.31%) and Spanish *hacer*-LVCs (53.62%). For this verb, all languages present percentages of determination of more than half of instances in my sample. With regard to bare nominals, both English and German have more bare plurals, even if to different degrees: 11.39% and 36% respectively, compared to bare singulars with only a 6.03% in English and 6.69% in German. In contrast, Romance languages have a higher presence of bare singulars: 15.96% in Catalan, and 28.19% in Spanish; than bare plurals with only 14.28% in Catalan and 18.19% in Spanish. This distribution is statistically significant when all languages are compared ($\chi2(6)= 998.98$, p<.0001, Cramer's V=0.25), as well as within language families: when the two Germanic languages are compared statistical tests prove a significant difference ($\chi2(2)=356.86$, p<.0001, Cramer's V=0.297, medium effect), and also between the two Romance languages ($\chi2(2)=109.79$, p<.0001, Cramer's V=0.166, small effect).
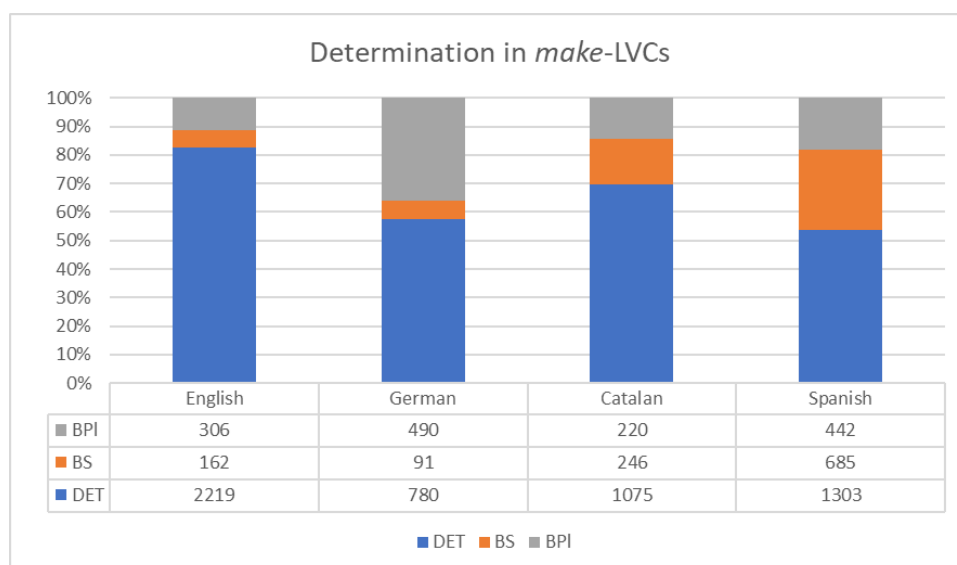
Figure 31. Distribution of the determination in make-*LVCs*

The distribution of compulsory and optional determiners in *make*-LVCs (Table 39) is especially explanatory in the case of English and Spanish. In English, the majority of LVCs have an obligatory determiner, while in Spanish most LVCs have an optional determiner. Therefore, differences in determination percentages could be explained in connection to the general possibilities to accept bare nominals of each language.

	Optional DET	Obligatory DET
make	Progress, profit, adjustment, mention	Decision, mistake, choice, call, claim, contribution, payment, announcement, assumption, accusation
machen	Hoffnung ('hope'), Anfang ('beginning'), Sport ('sport'), Eindruck ('impression'), Gedanke ('thought'), Sorge ('worry'), Angabe ('detail'), Erfahrung ('experience')	Fehler ('mistake'), Foto ('picture'), Ausführung ('implementation'), Beobachtung ('observation')
fer	Referència ('reference'), esforç ('effort'), feina ('work'), ús ('use'), anàlisi ('analysis'), dibuix ('drawing'), viatge ('trip'), visita ('visit'), fotografia ('picture')	Pregunta ('question'), petó ('kiss'), abraçada ('hug'), aclariment ('clarification')

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study

hacer	Carrera ('run'), énfasis ('emphasis'), broma ('joke'), colección ('collection'), esfuerzo ('effort'), referencia ('reference'), uso ('use'), ejercicio ('exercise'), análisis ('analysis'), investigación ('research'), declaración ('declaration')	Pregunta ('question'), viaje ('trip')

Table 39. Distribution of obligation and optionality of a determiner in *make*-LVCs

These results are aligned with the different tendencies of both languages in the nominal domain, with English prioritizing DPs in the object position and Romance languages being more flexible in accepting bare nominals in that position, although the results for German do not correspond to the general Germanic tendencies (Chierchia, 1998).

Figure 32 shows that English *make*-LVCs select nominals introduced by a determiner in all lemmas except for *make progress* which has a predominance of bare singulars (61.94%), versus only a 37.42% of determined nominals. In this case, the fact that the noun *progress* is a mass noun is key to understand its divergent behaviour from the rest of lemmas.

	decision	mistake	choice	call	progress	claim	contribution	profit	payment	adjustment	announcement	assumption	accusation	mention
BPl	19	5	21	26	1	9	53	10	35	23	23	10	68	3
BS	0	0	0	1	96	1	1	36	5	0	1	0	0	21
DET	181	194	175	171	58	186	134	153	118	177	176	190	130	176
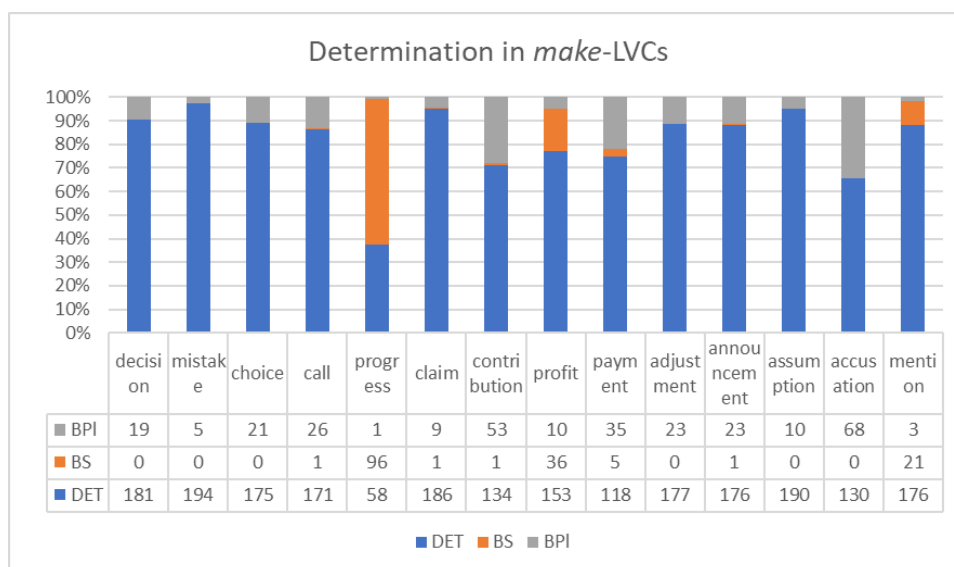
Figure 32. Distribution of the determination in English make-*LVCs per lemma*

Figure 33 shows that German *machen*-LVCs also combine predominantly with determined nominals.
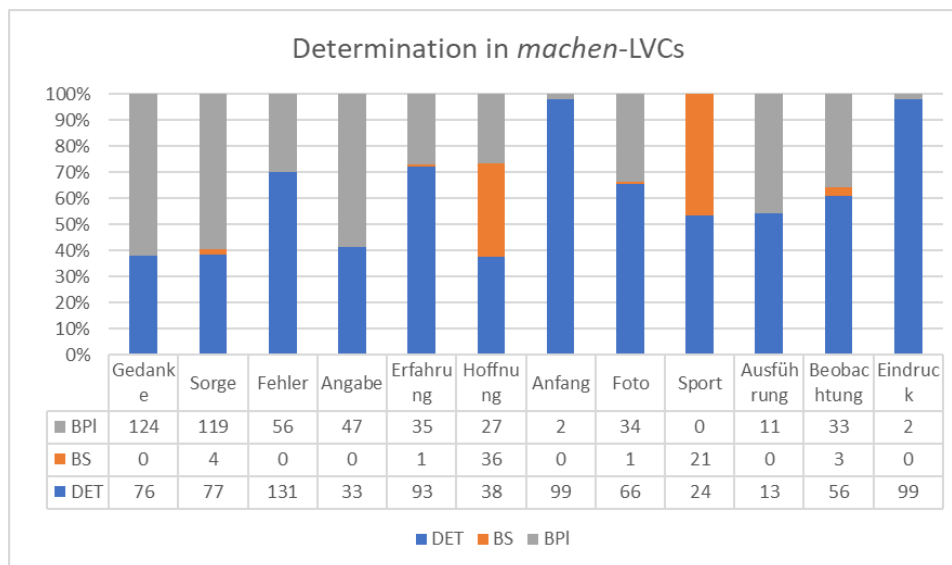
Figure 33. Distribution of the determination in German machen*-LVCs per lemma*

This LV, however, shows a tendency to select bare plurals more often than bare singulars. Unlike *nehmen* (Figure 18) and *geben* (Figure 3), only two lemmas in *machen*-LVCs coappear with bare singulars in high proportions: 35.64% of bare singulars in *Hoffnung machen,* and 46.67% of bare singulars in *Sport machen*, as in (113).

(113)   a. Aber es gibt in dieser Krise auch viele Entwicklungen, die *Hoffnung machen*, die Perspektive geben.   (DWDS, 2020)

'But there are also many developments in this crisis, which give hope and perspective'

b. Vielleicht sollte sie wirklich bald mal aufhören zu rauchen und *Sport machen*,   (DWDS, 2003)

'Maybe she should stop smoking really soon and do sport'

However, most high frequent LVCs select bare plurals more often: 64% of bare plurals in *Gedanke(n) machen*, 59.50% of bare plurals in *Sorge(n) machen*, 58.75% of bare plurals in *Angabe(n) machen,* but only 25.13% of bare plurals in *Fehler machen.* The fact that *Gedanke(n), Sorge(n)* and *Angabe(n)* usually coappear with plurals (regardless of whether they are introduced by a determiner or not), as in (114), points towards a certain degree of lexicalization of these constructions, which will be discussed when modification is taken into account (section §3.6.2 below).

(114) a. Ihr *macht* euch sogar *Gedanken* über den Stand der Christianisierung in Island,   (DWDS, 2006)

'You even reflect on the state of Christianization in Iceland'

b. Ich *mache* mir *Sorgen*.   (DWDS, 2006)

'I worry'

c. Dank der exzellenten Arbeit unserer Streitkräfte haben Dorfbewohner *Angaben gemacht*, wo die Rebellen versteckt sein könnten.

(DWDS, 2011)

'Thanks to the excellent work of our forces the inhabitants have given details of where the rebels could be hiding'

Figure 34 shows that the majority of *fer*-LVCs select nominals introduced by a determiner, although the determiner is not obligatory in the majority of lemmas (see Table 29 above).

The only exceptions are some of the most frequent LVCs: *fer referència* which selects bare singulars in 80.30% of instances, *fer ús* in 48.51%, and *fer feina* in 15%, as illustrated in (115).

Determination in *fer*-LVCs

	referència	esforç	feina	ús	pregunta	visita	viatge	petó	anàlisi	dibuix	fotografia	abraçada	aclariment
BPI	3	34	16	1	44	18	17	32	9	21	22	1	2
BS	159	0	30	49	1	6	0	0	1	0	0	0	0
DET	36	166	154	51	86	75	139	168	95	42	41	15	7
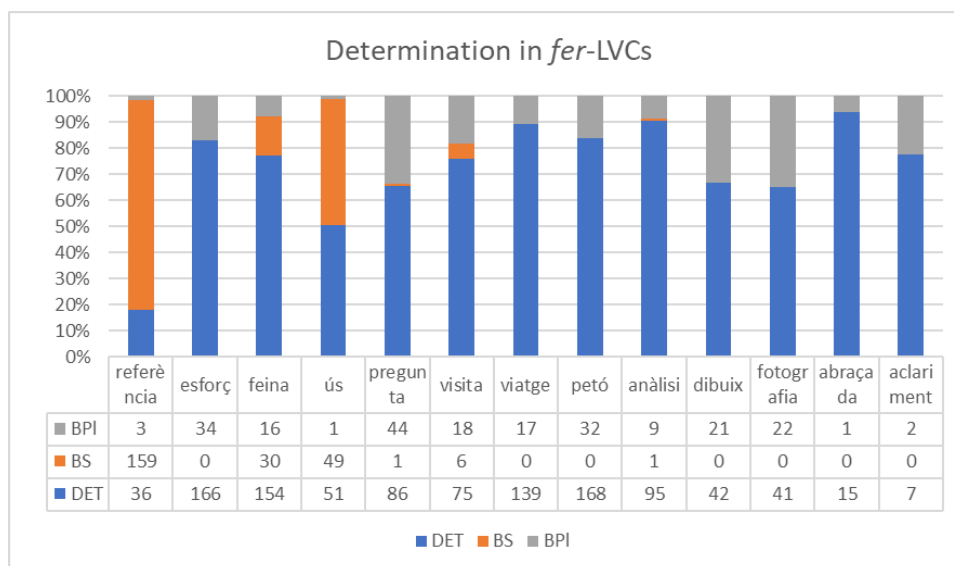
Figure 34. Distribution of the determination in Catalan fer-*LVCs per lemma*

(115) a. *Feia referència* a un llenguatge que és anàleg al que parla de la divinitat:

(CTILC, 2003)

'It made reference to a language which is analogous to [the language] that speaks about the divinity'

b. Jo no en puc *fer ús* perquè ja no tinc esma per viatjar i aquests indrets queden lluny d'Amsterdam. (CTILC, 2000)

'I cannot use it because I don't have the energy to travel, and these places are far from Amsterdam'

c. Era com la precisió i l'exactesa que exigeix *fer feina* amb la llengua. (CTILC, 2001)

'It was like the precision and accuracy that working with language requires'

In contrast, Figure 35 shows that the LV *hacer* in Spanish LVCs has a more prevalence of combination with bare nominals.
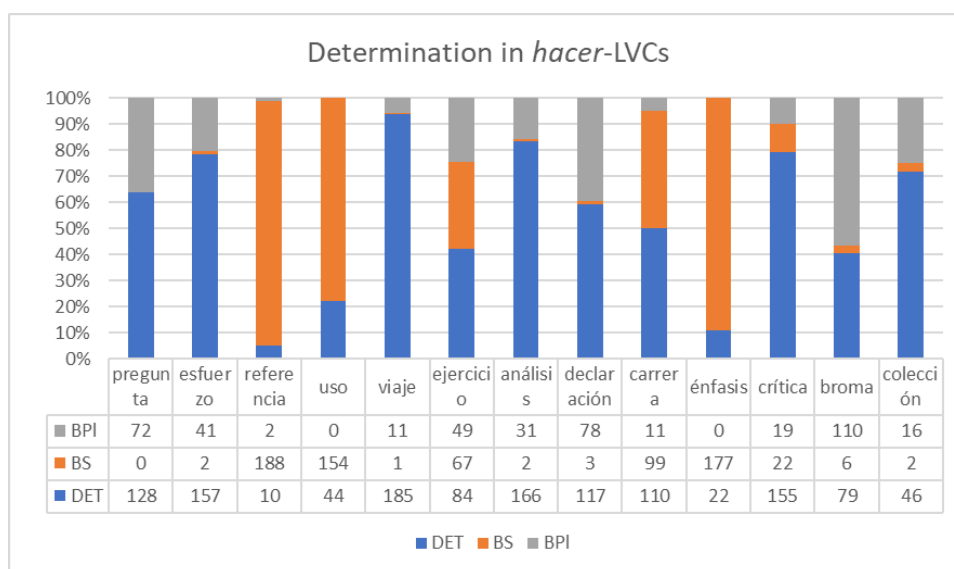
Figure 35. Distribution of the determination in Spanish hacer-*LVCs per lemma*

The frequency of bare singulars is most especially found in some lemmas: *hacer referencia* (94%)*, hacer énfasis* (88.94%), *hacer uso* (77.78%)*, but also others like *hacer carrera* (49.50%), or *hacer ejercicio* (33.50%), as illustrated in (116)*. This contrasts with the tendency observed for Catalan in Figure 34, although some of the lemmas with predominant bare singulars are shared (i.e. *referencia/referència* 'reference' and *uso/ús* 'use').

(116)   a. los cuales *hizo referencia* en el capítulo anterior. (CORPES XXI, 2001)

'to which he made reference in the previous chapter'

b. se propone *hacer uso* de los denominados "árboles de decisión".

(CORPES XXI, 2002)

'Using the so-called decision trees was proposed'

163

c. Es vital que su hijo *haga ejercicio*. (CORPES XXI, 2002)

'It is essential that his son does exercise'

d. El enfoque "alternativo", por su parte*, hace énfasis* en la competencia

como estrategia, (CORPES XXI, 2001)

'The alternative focus, specifically, emphasizes the competetition as a

strategy'

When determination and modification are compared in bare nominals, Figure 36 shows that English bare nominals are modified in 37.04% of bare singulars and 72.22% of bare plurals, the differences of which are statistically significant ($\chi2$ (1)=54.66, p<.0001, Cramer's V=0.342, medium effect). In the case of German bare nominals in *machen*-LVCs, these are modified in 23.08% of bare singulars and 34.90% of bare plurals. The differences in this case are only slightly statistically significant ($\chi2$ (1)=4.85 p=0.02, Cramer's V=0.091, negligible effect). In the case of Catalan, *fer*-LVCs have 21.95% of bare singulars modified and 40% of bare plurals, and the differences are also statistically significant ($\chi2$ (1)=17.86, p<.0001, Cramer's V=0.196, small effect). In English, German and Catalan a similar trend is found, thus, with bare plurals being more often modified than bare singulars, while the case of Spanish is the opposite, which is also not consistent with the results from the two previous LVs. The results for Spanish *hacer*-LVCs are the following: 52.99% of bare singulars are modified vs. 37.10% of bare plurals, and their difference is also significant ($\chi2$ (1)=27.24, p<.0001, Cramer's V = 0.155, small effect).

	English BS	English BPl	German BS	German BPl	Catalan BS	Catalan BPl	Spanish BS	Spanish BPl
UNMOD	102	85	70	319	192	132	322	278
MOD	60	221	21	171	54	88	363	164
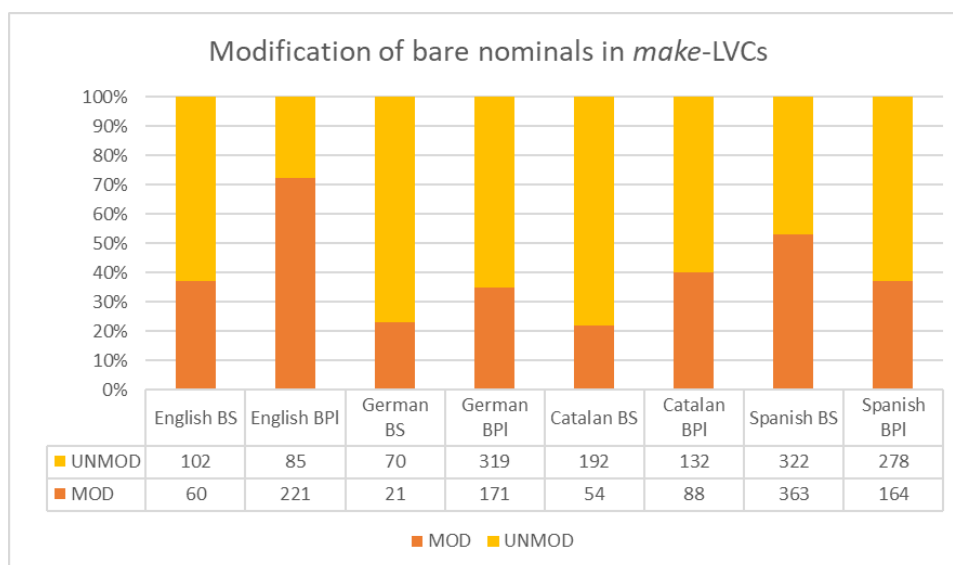
Figure 36. Distribution of the modification of bare nominals in make-*LVCs*

After this first correlation between determination and modification, the next section will present an analysis of the different subclasses of modification that have been included with a special focus on adjectival modification.

3.6.2. Modification in *make*-LVCs

The results on adjectival modification in *make*-LVCs show that the most frequently modified nominals within these constructions are found in English (60.85%), which contrasts with the results in the rest of the languages, all below half of the instances: 45.19% in Spanish, 37.25% in Catalan, and 28.36% in German (Figure 37).
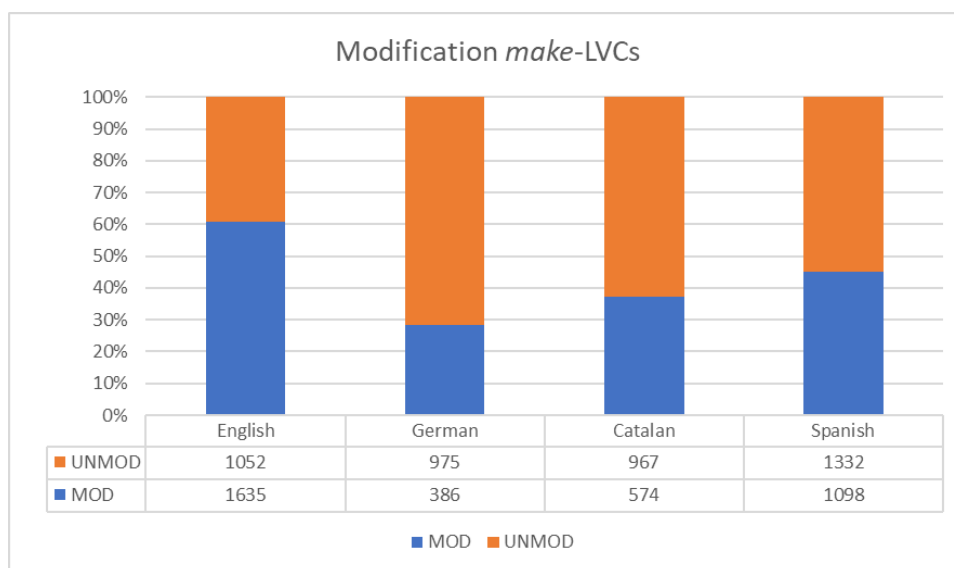
Figure 37. Distribution of modification in make-*LVCs*

This distribution is statistically significant when all languages are compared ($\chi 2(3)$=457.04, p<.0001, Cramer's V=0.239, medium effect), and within language families: when Germanic languages are compared statistical tests prove a significant difference ($\chi 2(1)$=381.39, p<.0001, Cramer's V=0.307, medium effect), and also do Romance languages ($\chi 2(1)$=24.37, p<.0001, Cramer's V = 0.078, negligible effect).

Table 40 shows that adjectival modification is the most common in three languages: 38.89% in English (versus 21.96% of the rest of modification subclasses), 26.38% in German (versus 1.98% of the rest of modification subclasses), and 23.04% in Catalan (versus 14.21% of the rest of modification subclasses). In contrast, PP modification is more prominent (22.84%) than adjectival modification (17.65%) in Spanish *hacer*-LVCs.

Modification	English	German	Catalan	Spanish
Unmod (No)	1052 (39.15%)	975 (71.64%)	967 (62.75%)	1332 (54.81%)
Adj	1045 (38.89%)	359 (26.38%)	355 (23.04%)	429 (17.65%)
PP	227 (8.45%)	7 (0.51%)	139 (9.02%)	555 (22.84%)
Rel	105 (3.91%)	15 (1.10%)	19 (1.23%)	16 (0.66%)
NC	140 (5.21%)	-	-	-
Gen	2 (0.07%)	1 (0.08%)	-	-
Various	116 (4.32%)	4 (0.29%)	61 (3.96%)	98 (4.04%)
TOTAL	2687	1361	1541	2430

Table 40. Distribution of modification subclasses in *make*-LVCs

Regarding PP modification, this subclass shows a considerable number of results in Catalan (9.02%) and in English (8.45%), as illustrated in (117). Moreover, Rel and Gen have residual results (3% or below); while NC is only found in English (5.21%), as in (118), and Gen are only found in Germanic languages. The differences in the choice of Germanic languages for NC and Gen and Romance languages for PP reflect the general modification tendencies of these language families, with the exception of English *make*-LVCs.

(117)   a. tenía por fin que hacer *una crítica de los consultorios sentimentales* de los periódicos de la época;   (CORPES XXI, 2003)
'he had to finally make a review of the problem page in the newspaper of the period'
b. i perquè ens fessin *preguntes de tota mena,*   (CTILC, 2005)
'and so that they could make questions of any kind'
c. The preliminary reports make *no mention of such work.* (COCA, 2019)

(118)   And then Vince made *a phone call*,   (COCA, 2018)

A closer look into these divergences will be analysed through the examination of the degree of frequency of LVCs (§3.6.2.1), as well as the study of each individual LVC in every language (§3.6.2.2, §3.6.2.3, §3.5.2.4 and §3.5.2.5).

3.6.2.1. Degree of frequency

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 corroborate whether the degree of frequency influences the overall results.

Figure 38 shows that the three groups of frequency in English *make*-LVCs present close results with an increasing tendency for modified nominals: high frequent LVCs in 55.06%, medium-frequency in 61.40%, and low-frequency in 65.93%. The differences are statistically significant ($\chi 2$ (2)=24.24, p<.0001, Cramer's V=0.095, small effect).
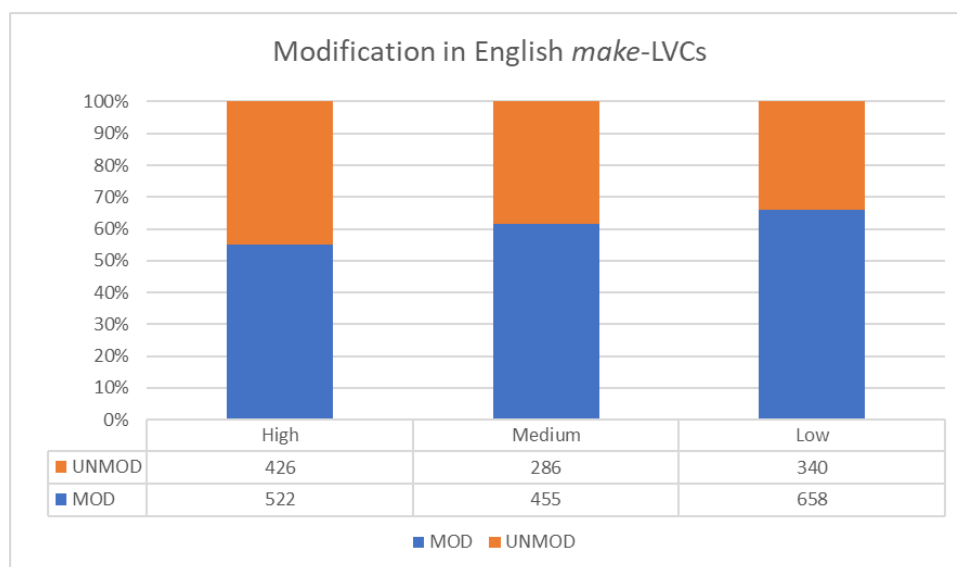
Figure 38. Distribution of modification in English make-*LVCs according to the degree of frequency (high, medium or low)*

The results for English point to an inverse correlation between degree of frequency and levels of modification: the more frequent a LVC is, the less often it appears modified. The kind of modification plays a role in this LV for English, as seen in Table 41. In this case, adjectival modification is the most predominant modification subclass in all three groups of frequency: 39.87% in high frequency, 50.74% in medium frequency, and 29.16% in low frequency. But only the low-frequency LVCs show relevant preference for both PP modification (18.94%) and relative clauses (7.71%), as well; while noun classifiers are 8.02% of instances in the high frequency group and 5.40% in the medium frequency.

Modification	High	Medium	Low
Unmod (No)	426 (44.94%)	286 (38.60%)	340 (34.07%)
Adj	378 (39.87%)	376 (50.74%)	291 (29.16%)
PP	31 (3.27%)	7 (0.95%)	189 (18.94%)
Rel	15 (1.58%)	13 (1.75%)	77 (7.71%)
NC	76 (8.02%)	40 (5.40%)	24 (2.41%)
Gen	-	2 (0.27%)	-
Various	22 (2.32%)	17 (2.29%)	77 (7.71%)
TOTAL	948	741	998

Table 41. Distribution of modification subclasses in *make*-LVCs according to the degree of frequency

Figure 39 shows that the trend is the same for German *machen*-LVCs, but with a lower proportion of modification. Again, the low-frequent LVCs are more often modified (43.25%) than the other two groups, followed by the results of medium-frequent LVCs (35.56%), whereas the high-frequency LVCs show the lowest percentage of modification (16.64%). The statistics support the significance of the difference between the degree frequency groups ($\chi2$ (2)= 93.35, p<.0001, Cramer's V = 0.262, medium effect).
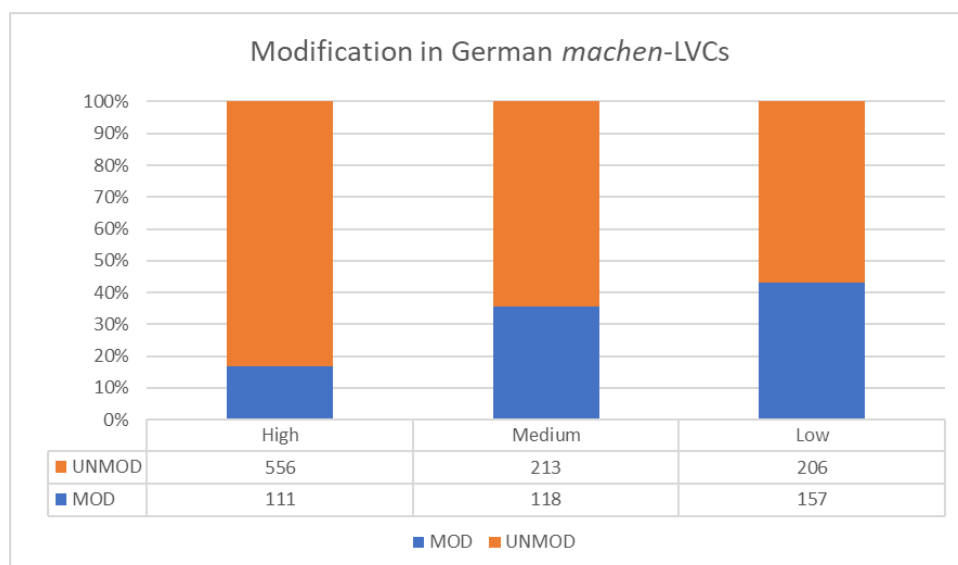
Figure 39. Distribution of modification in German machen-*LVCs according to the degree of frequency (high, medium or low)*

Table 42 shows that the only predominant subclass of modification in the case of *machen*-LVCs is adjectival modification: 40.77% in the low-frequency group, 34.14% in medium-frequency, and 14.69% in low-frequency *machen*-LVCs. The rest of subclasses are residual (3% or below).

Modification	High	Medium	Low
Unmod (No)	556 (83.36%)	213 (64.35%)	206 (56.75%)
Adj	98 (14.69%)	113 (34.14%)	148 (40.77%)
PP	6 (0.90%)	1 (0.30%)	-
Rel	4 (0.60%)	3 (0.91%)	8 (2.20%)
NC	-	-	-
Gen	-	1 (0.30%)	-
Various	3 (0.45%)	-	1 (0.28%)
TOTAL	667	331	363

Table 42. Distribution of modification subclasses in *machen*-LVCs according to the degree of frequency

Figure 40 shows that the distribution of modification in Catalan *fer*-LVCs is more predominant with low-frequency LVCs, with more than half of instances (52.73%), which contrast with medium-frequency (36.68%) and high-frequency LVCs (31.27%). The difference in this distribution is also statistically significant ($\chi 2$ (2)=35.5, p<.0001, Cramer's V = 0.152, small effect). In this case, the percentual divergence can be explained by the lower number of examples extracted for the low-frequency LVCs, due to the short number of examples available in the CTILC corpus for the 21st Century. Also, this tendency is consistent with the previous LV *donar* ('give', Figure 10 above), where low-frequency LVCs are always more often modified than the other two groups.
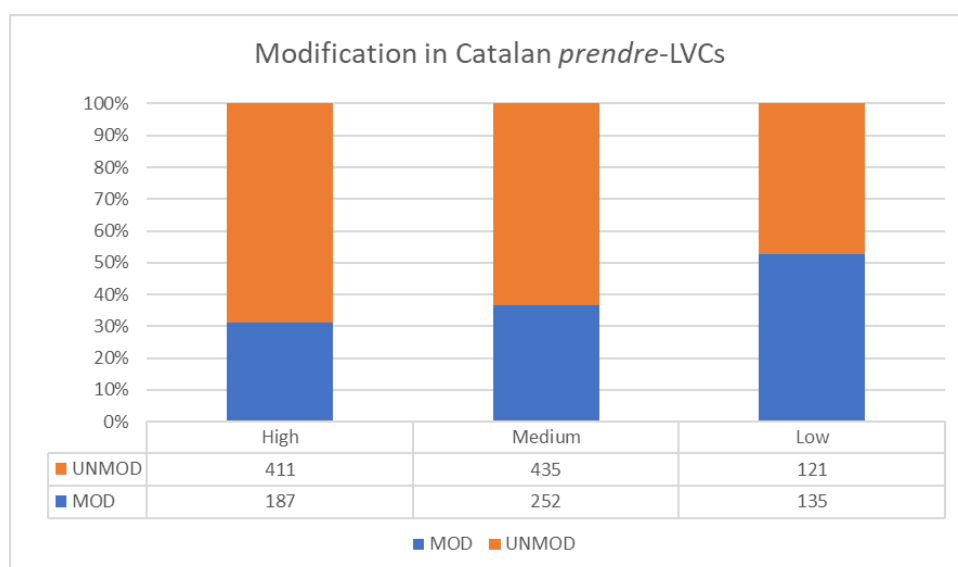
Figure 40. Distribution of modification in Catalan fer-*LVCs according to the degree of frequency (high, medium or low)*

Table 43 confirms the tendency of adjectival modification to be predominant in all three groups of frequency: 30.47% in low frequency, 22.24% in high frequency, and 20.96% in medium frequency. Regarding the rest of subclasses, only PP modification stands out with higher prominence in the low-frequency group (12.11%), followed by medium frequency (10.04%) and high frequency to a lesser extent (6.52%). The rest of subclasses are residual for this language.

Modification	High	Medium	Low
Unmod (No)	411 (68.73%)	435 (63.32%)	121 (47.27%)
Adj	133 (22.24%)	144 (20.96%)	78 (30.47%)
PP	39 (6.52%)	69 (10.04%)	31 (12.11%)
Rel	5 (0.84%)	9 (1.31%)	5 (1.95%)
NC	-	-	-
Gen	-	-	-
Various	10 (1.67%)	30 (4.37%)	21 (8.20%)
TOTAL	598	687	256

Table 43. Distribution of modification subclasses in *fer*-LVCs according to the degree of frequency

For Spanish *hacer*-LVCs, Figure 41 shows that the difference between the three frequency groups is low: medium-frequency LVCs (40.68%) and high-frequency LVCs (43.73%) have a similar percentage of modified instances, and low-frequency LVCs have half of instances modified (50.84%). These differences, despite being perceptively minimal, do show statistical significance ($\chi 2$ (2)=17.99, p=0.0001, Cramer's V=0.086, small effect).
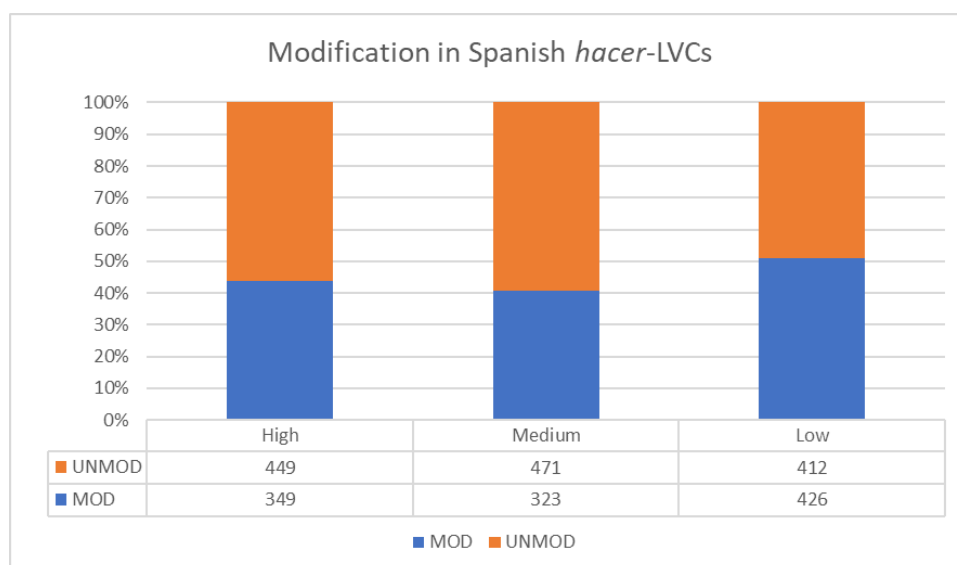
Figure 41. Distribution of modification in Spanish hacer-LVCs according to the degree of frequency (high, medium or low)

Table 44 shows two subclasses of modification as the most predominant: PP modification and adjectival modification. Two groups prefer the former: 20.80% in high-frequency LVCs (versus 18.17% of adjectival modification) and 30.31% in low-frequency (versus 16.23% of adjectival modification; while medium-frequency LVCs prefer adjectival modification in 18.63%, closely followed by PP modification in 17% of instances. The rest of subclasses are residual for this language.

Modification	High	Medium	Low
Unmod (No)	449 (56.27%)	471 (59.32%)	412 (49.16%)
Adj	145 (18.17%)	148 (18.64%)	136 (16.23%)
PP	166 (20.80%)	135 (17%)	254 (30.31%)
Rel	5 (0.63%)	6 (0.76%)	5 (0.60%)
NC	-	-	-
Gen	-	-	-
Various	33 (4.13%)	34 (4.28%)	31 (3.70%)
TOTAL	798	794	838

Table 44. Distribution of modification subclasses in *hacer*-LVCs according to the degree of frequency

The results for the distribution of modification present an increasing tendency of modification from high-frequency to low-frequency LVCs, where the latter has a higher proportion of modification. This tendency is found in all four languages, although with different proportions. Regarding the modification subclasses, adjectival modification is

171

the most prominent in all languages, but PP modification is also relevant in some groups in the two Romance languages. Interestingly, PP modification is highlighted in low-frequency *make*-LVCs in English too.

In order to analyse the possible reasons for this shared tendency in all four languages, each selected LVC will be analysed in the subsequent sections.

3.6.2.2. English make-LVCs

An examination of each English LVC in my sample shows that the LVCs with the highest percentage of modification are in the high-frequency group (Figure 42): *make a decision* (80%) and *make a call* (74.74%), while the other lemmas in the same group have much lower percentages of modification: *make a mistake* (35.68%), *make a choice* (47.96%) and *make progress* (31.61%). In the medium-frequency group the results are more systematic, since three out of four have the majority of instances modified: *make a claim* (62.76%), *make a contribution* (69.15%), *make profit* (58.79%), and *make payment* (53.80%). Finally, low-frequency LVCs present percentages above 50% in all lemmas: *make adjustment* (54.50%), *make announcement* (71.50%), *make assumption* (61.50%), *make accusation* (50.51%) and *make mention* (91.50%).
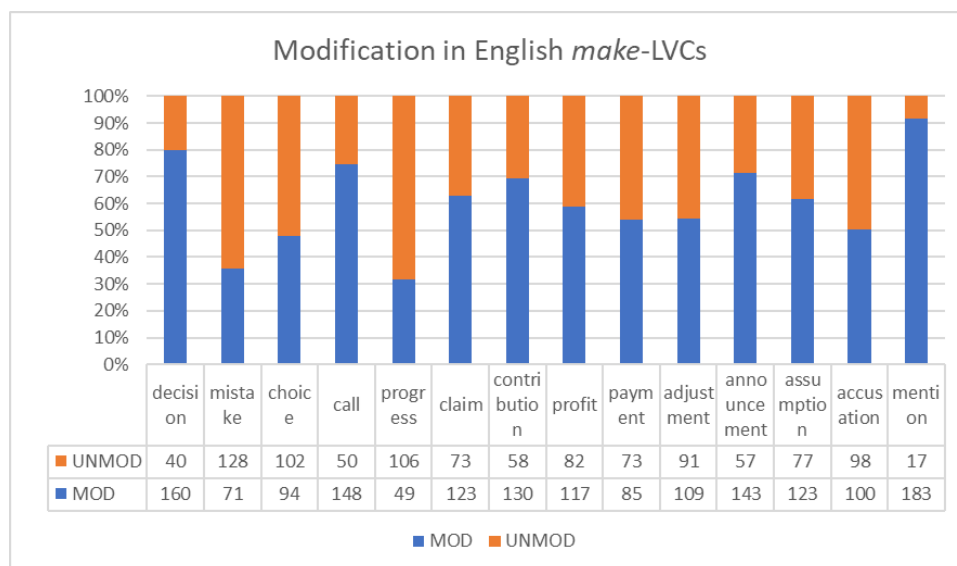
Figure 42. Distribution of modification in English make-*LVCs per lemma*

Table 45 shows that adjectival modification is predominant in all lemmas except for *make mention* (72.50% of PP modification versus 4.50% of adjectival modification), which can be due to the selection of an *of*-PP argument (that is, *make mention of*).

Modification	decision	mistake	choice	call	progress	claim	contribution	profit	payment	adjustment	assumption	announcement	accusation	mention
Unmod (No)	40 (20%)	128 (64.32%)	102 (52.04%)	50 (25.26%)	106 (68.39%)	73 (37.24%)	58 (30.85%)	82 (41.21%)	73 (46.20%)	91 (45.50%)	77 (38.50%)	57 (28.50%)	98 (49.49%)	17 (8.50%)
Adj	146 (73%)	39 (19.60%)	75 (38.27%)	74 (37.37%)	44 (28.38%)	82 (41.84%)	128 (68.09%)	111 (55.78%)	55 (34.81%)	98 (49%)	27 (13.50%)	103 (51.50%)	54 (27.27%)	9 (4.50%)
PP	-	23 (11.56%)	8 (4.08%)	-	-	5 (2.56%)	-	-	2 (1.26%)	-	20 (10%)	11 (5.50%)	13 (6.57%)	145 (72.50%)
Rel	-	6 (3.01%)	5 (2.55%)	-	4 (2.58%)	11 (5.61%)	-	-	2 (1.26%)	2 (1%)	52 (26%)	7 (3.50%)	14 (7.07%)	2 (1%)
NC	2 (1%)	-	-	74 (37.37%)	-	12 (6.12%)	-	5 (2.51%)	23 (14.56%)	9 (4.5%)	1 (0.5%)	14 (7%)	-	-
Gen	-	-	-	-	-	-	-	-	2 (1.26%)	-	-	-	-	-
Various	12 (6%)	3 (1.51%)	6 (3.06%)	-	1 (0.65%)	13 (6.63%)	2 (1.06%)	1 (0.5%)	1 (0.65%)	-	23 (11.50%)	8 (4%)	19 (9.60%)	27 (13.50%)
TOTAL	200	199	196	198	155	196	188	199	158	200	200	200	198	200

Table 45. Distribution of modification subclasses in *make*-LVCs per lemma

Other lemmas also show relevant presence of PP modification: *make mistake* (11.56%), *make assumption* (10%), *make accusation* (6.57%), *make announcement* (5.50%), and *make choice* (4.08%), as in (119). In *make*-LVCs there are other two subclasses of modification which show relevant results: relative clauses and noun classifiers. The former is present in the following lemmas: *make assumption* (26%), *make accusation* (7.07%), and *make claim* (5.61%), as in (120); the latter is found in *make call* (37.37%), *make payment* (14.50%), *make claim* (6.12%) and *make adjustment* (4.50%), as in (121).

(119)   a. The proposal makes *no mention of the United States*.   (COCA, 2019)

b. The same thing happens to consumers who make *the mistake of engaging with callers*,   (COCA, 2018)

c. According to the Mental Capacity Act 2005, we make *an assumption of capacity*,   (COCA, 2011)

d. that some Afghan candidates have made *accusations of technical problems* relating to the ink applied to voters' thumbs.   (COCA, 2004)

e. He has made *announcements about the plans* for his business interests   (COCA, 2016)

f. In reality it doesn't work because people make *choices about where they want to live*.   (COCA, 2002)

(120)   a. I make *the assumption that they must employ* at least one such person,

(COCA, 2012)

b. On social media, many fans are upset that Star made *accusations that he never publicly backed up*. (COCA, 2019)

c. He can make *the claim that it was not unreasonable to be optimistic*,

(COCA, 2007)

(121)   a. Will you please excuse me while I make *an emergency phone call* totally unrelated to this? (COCA, 2014)

b. You fail to make *your mortgage payment* by the due date.

(COCA, 2012)

c. they have a right to make *an asylum claim* and get protection if they need it. (COCA, 2018)

d. he will again make *the language adjustment* based on his environment.

(COCA, 2009)

The presence of these three subclasses of modification in *make*-LVCs proves to be more relevant than in the two previously analysed English LVCs (with *give* and *take*). The comparison with the results in the rest of the languages under analysis will be presented in the discussion section (§3.6.3), as it is of relevance.

The distribution of the adjectival modification in English *make*-LVCs presents an irregular distribution per lemma. If the focus is put on the LVCs with bigger percentages of adjectival modification, no lexicalization patterns are to be found. In the case of *make a decision*, there are five adjectives which are predominant (in 56/158 instances, 35.44%), as in (122): *big* (13/158, 8.22%), *good* (13/158, 8.22%), *informed* (10/158, 6.33%), *difficult* (10/158, 6.33%), and *quick* (10/158, 6.33%).

(122)   a. I can't make such a *big* decision so quickly!   (COCA, 2019)

b. we know what the heck we're deciding on and we make a *pretty good* decision. (COCA, 2015)

c. There is something for you as a consumer to make a *better informed* decision. (COCA, 2014)

d. I had to make a very *difficult* decision.   (COCA, 2014)

e. So he had to make a very *quick* decision.   (COCA, 2011)

In the case of *make an announcement,* the second most modified LVC, there are four predominant adjectives with which it co-occurs, as in (123): *big* (20/103, 19.42%), *public* (10/103, 9.71%), *official* (9/103, 8.74%) and *formal* (7/103, 6.80%).

(123)   a. North Carolina's governor, Roy Cooper, made a *big* announcement last fall.   (COCA, 2019)

b. On July 15, 2011, Marc and I made the *public* announcement that we were going to divorce. Hardest. Day. Ever.   (COCA, 2014)

c. So, we made an *official* announcement.   (COCA, 2013)

d. In all likelihood, I will. I have not made a *formal* announcement.   (COCA, 2004)

The third most modified LVC, *make a contribution*, follows a similar tendency with four predominant adjectives among a long list of adjectives which are not repeated. The most common adjectives constitute a 25.38% of all instances (33/130), as in (124): *significant* (12/130, 9.23%), *important* (8/130, 6.15%), *positive* (7/130, 5.38%), and *meaningful* (6/130, 4.62%).

(124)   a. A greenhouse large enough to make a *significant* contribution to supplying your family with homegrown food year-round should be at least 10 by 12.   (COCA, 2013)

b. It also argues that carbon capture and storage would " begin to make an *important* contribution by 2035   (COCA, 2012)

c. Fitzgerald claims we should accept and tolerate eccentrics as they frequently have *positive* contributions to make.   (COCA, 2012)

d. fields of study exist, especially in science, where an amateur can make *meaningful* contributions.   (COCA, 2004)

A similar trend is found in *make (a) profit* with five predominant adjectives (in 42/112 instances, 37.50%), as in (125): *big* (16/112, 14.29%), *quick* (7/112, 6.25%), *large* (7/112, 6.25%), *little* (6/112, 5.35%), and *healthy* (6/112, 5.35%).

(125)   a. although I'm sure they can make a *big* profit selling it to Halliburton to pump down a well bore.   (COCA, 2012)

b. that China is still a prime destination for poachers trying to make a *quick* profit.   (COCA, 2012)

c. it looked like a much easier way to make much *larger* profits.

(COCA, 2012)

d. They thought they might make a *little* profit on it if Ali's career went

well, but they knew that most boxers. (COCA, 2017)

e. Its stations make a *healthy* profit showing syndicated entertainment

programs when their competitors show the national evening newscasts.

(COCA, 2002)

In these cases, which are the most often modified *make*-LVCs, there is also a wide range of adjectives with which they co-occur in fewer cases. These are both relational (126) and qualitative (127) adjectives.

(126)   a. defeated Members of Congress coming back to Washington to make

their last *legislative* decisions. (COCA, 2012)

b. and it has not made any *official* announcement about Clarke since.

(COCA, 2017)

c. Although women-owned businesses make a staggering *economic*

contribution (COCA, 2000)

d. then it must be in these areas where it doesn't necessarily make a

financial profit to talk about it. (COCA, 2012)

(127)   a. To go home and make the *first mature* decision of your life by sitting

down to talk with Myra, turn to Page 112. (COCA, 2014)

b. And one month ago its multi-billionaire inventor Darrius Sayle made

his *astonishing* announcement. (COCA, 2006)

c. Therefore, the purpose of the present study is to make a *modest*

contribution to that area. (COCA, 2006)

d. who could otherwise make an *easy* profit by investing proceeds at the

generally higher rates on taxable investments. (COCA, 2009)

No differences are found in this regard between LVCs that optionally select bare nominals (i.e. *make profit*) and those LVCs that must appear with an obligatory determiner (i.e. *make a decision, an announcement, a contribution*).

3.6.2.3. German machen-LVCs

Figure 43 presents the results of modification for each German *machen*-LVC. In the analysis of the individual LVCs, the tendency observed in previous sections is

confirmed: low-frequent LVCs are more modified than those of the other two frequency groups. The three LVCs which present a majority of instances with modification are in the low-frequency group: *Ausfürung machen* 'make implementation' (54.17%), *Beobachtung machen* 'make observation' (61.96%) and *Eindruck machen* 'make impression' (62.38%), but the other two lemmas display much lower percentages: *Foto machen* 'make picture' (20.79%) and *Sport machen* 'make sport' (6.67%). In the medium-frequency LVCs, the percentages are lower than half of the instances but above the mean for this language (28.36%, as seen in Figure 37): *Erfahrung machen* 'make experience' (44.19%), *Hoffnung machen* 'make hope' (37.62%) and *Anfang machen* 'make beginning' (22.77%). Lastly, the least often modified LVCs are the most frequent in German: *Gedanke machen* 'make thought' (6.5%), *Sorge machen* 'make worry' (10%), *Fehler machen* 'make mistake' (16.65%), with the exception of *Angabe machen* 'make detail' which presents modification percentages above the mean for *machen*-LVCs (56.25%).

Figure 43. Distribution of modification in German machen-*LVCs per lemma*

Table 46 shows that adjectival modification is the most predominant in all lemmas, and the rest of subclasses are residual with two exceptions: PP modification in *Angabe machen* 'make detail' (7.50%) and relative clauses in *Beobachtung machen* 'make observation' (5.44%), as illustrated in (128).

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study

Modification	Gedanke	Sorge	Fehler	Angabe	Erfahrung	Hoffnung	Anfang	Foto	Sport	Ausführung	Beobachtung	Eindruck
Unmod (No)	187 (93.50%)	180 (90%)	154 (82.35%)	35 (43.75%)	72 (55.81%)	63 (62.38%)	78 (77.23%)	80 (79.21%)	42 (93.33%)	11 (45.83%)	35 (38.04%)	38 (37.62%)
Adj	13 (6.50%)	19 (9.50%)	30 (16.04%)	36 (45%)	54 (41.86%)	36 (35.64%)	23 (22.77%)	18 (17.82%)	3 (6.67%)	13 (54.17%)	52 (56.52%)	62 (61.39%)
PP	-	-	-	6 (7.50%)	-	1 (0.99%)	-	-	-	-	-	-
Rel	-	1 (0.50%)	3 (1.61%)	-	2 (1.55%)	1 (0.99%)	-	3 (2.97%)	-	-	5 (5.44%)	-
NC	-	-	-	-	-	-	-	-	-	-	-	-
Gen	-	-	-	-	1 (0.78%)	-	-	-	-	-	-	-
Various	-	-	-	3 (3.75%)	-	-	-	-	-	-	-	1 (0.99%)
TOTAL	199	200	187	80	129	101	101	101	45	24	92	101

Table 46. Distribution of modification subclasses in *machen*-LVCs per lemma

(128)　　a. Angaben über Lage und innerhalb der Kirche macht Einhard nicht.

(DWDS, 2013)

'Einhard does not give details about the situation and inside the Church'

b. Lemke hat *eine Beobachtung* gemacht, *die zeigt, wie wichtig es ist*,

(DWDS, 2011)

'Lemke made an observation that shows how important it is'

If the co-appearance with certain adjectives is examined, the high frequency *machen*-LVCs have very low percentages of adjectival modification, as well as determination (see Figure 33). Likewise, they appear mainly in plural, which points towards a certain degree of lexicalization of such construction as *Gedanken machen* and *Sorgen machen*.

When the focus is placed on the LVCs with higher percentages of adjectival modification, there are tendencies of co-occurring adjectives with some lemmas. For

instance, *Angabe machen* has two adjectives with several co-occurrences, as in (129): *falsch* ('false', in 13/39, 33.33%) and *detailliert* ('detailed', in 5/39, 12.82%).

(129)   a. Sie haben *falsche* Angaben gemacht.   (DWDS, 2013)

      'They gave wrong details'

      b. Der Unternehmer machte *detaillierte* Angaben über die Art und Weise

          (DWDS, 2001)

      'The businessman gave particular details about the method'

Similar tendencies are found in the LVCs with more than half of examples modified by an adjective. This is the case of *Eindruck machen* with the adjectives *gut* ('good', in 15/63, 23.81%) and *erst* ('first', in 5/63, 7.94%), as in (130). At the same time, a tendency to co-occur with adjectives with an adverbial base is found, as in (131).

(130)   a. Sie wollte freundlich sprechen und so einen *guten* Eindruck machen.

          (DWDS, 2002)

      'She wanted to speak in a friendly way and, thus, give a good impression'

      b. ich mein nur, dass ich einen fürchterlichen *ersten* Eindruck mache.

          (DWDS, 2010)

      'I only mean that I give a terrible first impression'

(131)   a. Ich muß Dir aber auch sagen, daß Dein Bild einen *unbegreiflichen* Eindruck macht.   (DWDS, 2004)

      'I must tell you that your picture gives an inexplicable impression'

      b. Er hatte bei den letzten Telefonaten einen ungewöhnlich *fröhlichen* Eindruck gemacht   (DWDS, 2002)

      'In the last phone call, he had given an unusually friendly impression'

      c. Auch ich muss heute einen *ordentlichen* Eindruck machen

          (DWDS, 2000)

      'Also, I must give today a proper impression'

In the case of *Erfahrung machen,* there are also two adjectives with relevant presence in the adjectival modification: *gut* ('good', in 12/54, 22.22%) and *schlecht* ('bad', in 9/54, 16.67%), as in (132).

(132)   a. Wo dies geschieht, hat man *gute* Erfahrungen gemacht.

          (DWDS, 2000)

179

'Where this happens, one has good experiences'

b. hätte er damit nicht einmal *schlechte* Erfahrungen gemacht.

(DWDS, 2004)

'If he had not once bad experiences with this'

In *Beobachtung machen* the adjectival modification corresponds to mainly three adjectives (in 28/52, 53.85%), as in (133): *interessant* ('interesting, in 12/52, 23.08%), *verdächtig* ('suspicious', in 9/52, 17.31%) and *ähnlich* ('similar', in 7/52, 13.46%).

(133)   a. hat das polnische Gastronomenpaar eine *interessante* Beobachtung
           gemacht   (DWDS, 2014)
           'The Polish restaurateurs made an interesting observation'
           b. Die Polizei hofft nun auf Zeugen, die *verdächtige* Beobachtungen
           gemacht haben.   (DWDS, 2014)
           'The police hoped for signs, that made the suspicious observations'
           c. andere Leute hätten *ähnliche* Beobachtungen gemacht, erzählt sie.

(DWDS, 2013)

'other people had made similar observations, she explained'

Finally, in the case of *Ausführung machen* there is only one adjective which is repeated: *lang* ('long', in 3/13, 23.08%), as in (134). However, the number of modified instances is only 13, and the variety of adjectives found is wide, because the rest appear only in one instance and are not repeated.

(134)   dazu will ich jetzt keine *langen* Ausführungen machen. (DWDS, 2013)
           'I don't want to make a long implementation of this.'

In fact, adjectival modification results in *machen*-LVCs should be taken cautiously, since there is not enough proof that there are signs of lexicalization in the co-occurrence of certain nominals with specific adjectives. Low number of modified instances results in high percentages of co-occurrence with certain adjectives, although the number of examples is not higher than 15. Comparatively, the percentages are higher than in English *make*-LVCs because the number of examples is lower, but a closer look into the numbers shows that the raw numbers are similar.

3.6.2.4. Catalan fer-LVCs

Catalan *fer*-LVCs have shown a greater tendency to be modified in low-frequency LVCs (Figure 40 above). However, these results should be taken cautiously due to the low number of instances extracted for this group of LVCs. Hence, a need for an analysis of each LVC is especially relevant for this language.

In Figure 44, the high-frequent LVCs differ in their co-occurrence with modification: *fer referència* (13.13%), *fer esforç* (47.50%) and *fer feina* (33%). A similar tendency is found in medium-frequent LVCs, such as *fer ús* (80.20%), *fer pregunta* (35.11%), *fer visita* (39.39%), *fer viatge* (35.26%) and *fer petó* (15.50%). Finally, the low-frequency lemmas show higher percentages, but they represent fewer instances, with the exception of *fer anàlisi* (75.24%). The rest of the LVCs in this group are below the 100 instances: *fer dibuix* (46.03% in 63 instances), *fer fotografia* (23.81% in 63 instances), *fer abraçada* (50% in 16 instances) and *fer aclariment* (44.44% in 9 instances).

	referè ncia	esforç	feina	ús	pregu nta	visita	viatge	petó	anàlisi	dibuix	fotogr afia	abraça da	aclari ment
UNMOD	172	105	134	20	85	60	101	169	26	34	48	8	5
MOD	26	95	66	81	46	39	55	31	79	29	15	8	4

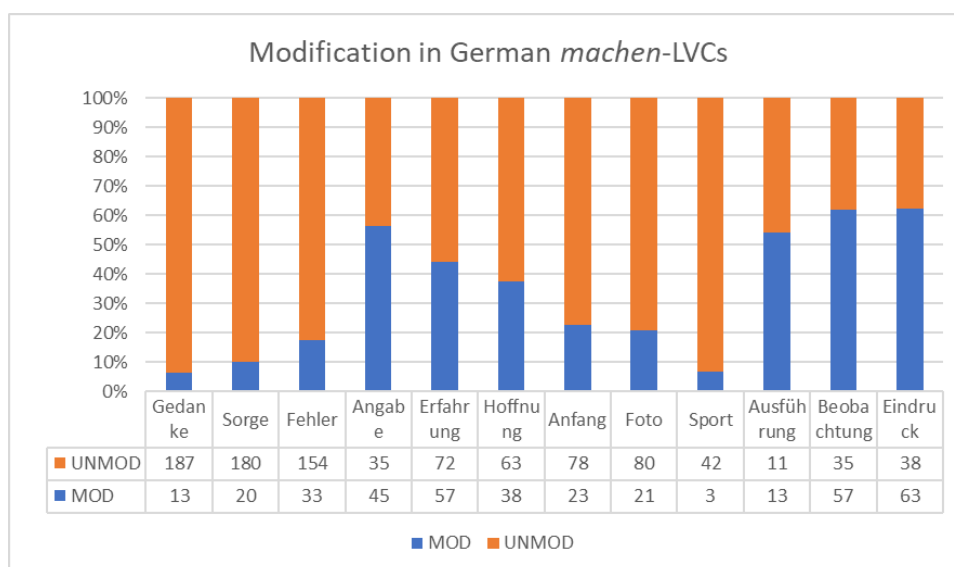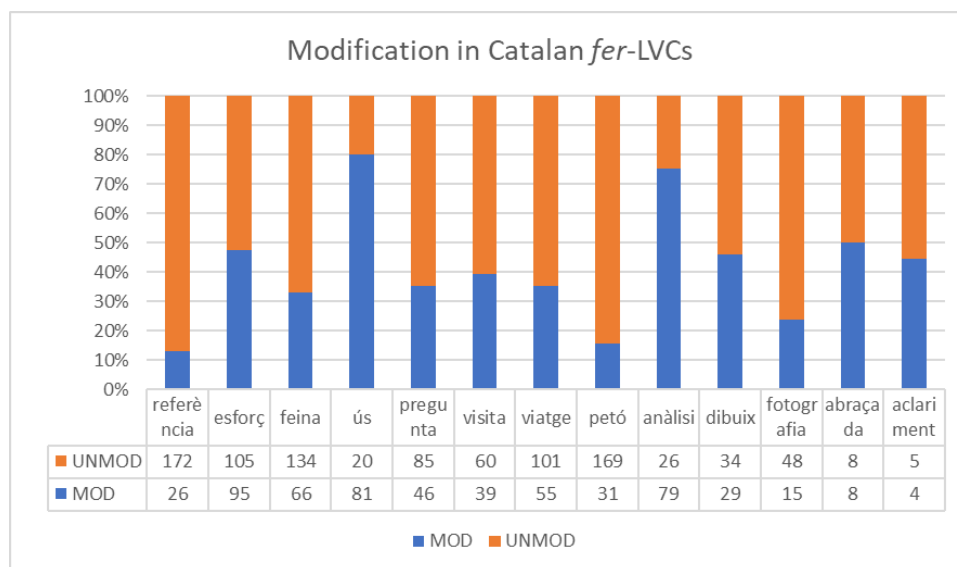*Figure 44. Distribution of modification in Catalan* fer-*LVCs per lemma*

Table 47 shows that adjectival modification is predominant in all lemmas except for *fer ús* (35.65% of PP modification versus 15.84% of adjectival modification).

Modification	referència	esforç	feina	ús	pregunta	visita	viatge	petó	anàlisi	dibuix	fotografia	abraçada	aclariment
Unmod (No)	172 (86.87%)	105 (52.50%)	134 (67%)	20 (19.80%)	85 (64.89%)	60 (60.61%)	101 (64.74%)	169 (84.50%)	26 (24.76%)	34 (53.97%)	48 (76.19%)	8 (50%)	5 (55.56%)
Adj	25 (12.63%)	70 (35%)	38 (19%)	16 (15.84%)	28 (21.37%)	32 (32.32%)	45 (28.85%)	23 (11.50%)	39 (37.14%)	21 (33.33%)	7 (11.11%)	7 (43.75%)	4 (44.44%)
PP	-	14 (7%)	25 (12.50%)	36 (35.65%)	11 (8.40%)	6 (6.06%)	10 (6.41%)	6 (3%)	18 (17.14%)	6 (9.52%)	6 (9.52%)	1 (6.25%)	-
Rel	1 (0.50%)	2 (1%)	2 (1%)	-	6 (4.58%)	1 (1.01%)	-	2 (1%)	2 (1.91%)	2 (3.18%)	1 (1.59%)	-	-
NC	-	-	-	-	-	-	-	-	-	-	-	-	-
Gen	-	-	-	-	-	-	-	-	-	-	-	-	-
Various	-	9 (4.50%)	1 (0.50%)	29 (28.71%)	1 (0.76%)	-	-	-	20 (19.05%)	-	1 (1.59%)	-	-
TOTAL	198	200	200	101	131	99	156	200	105	63	63	16	9

Table 47. Distribution of modification subclasses in *fer*-LVCs per lemma

Also, most LVCs show a relevance presence of PP modification: *fer esforç* (7%), *fer feina* (12.50%), *fer pregunta* (8.40%), *fer visita* (6.06%), *fer viatge* (6.41%), *fer anàlisi* (17.14%), *fer dibuix* (9.52%), *fer fotografia* (9.52%), and *fer abraçada* (6.25%, in only one instance), see (135). There is also one case of relevant distribution of relative clause modification: *fer pregunta* (4.58%), as in (136).

(135)   a. En principi, també podríem fer *ús d'expressions* usualment classificades com a noms propis.   (CTILC, 2001)

'In principle, we could also make us of expressions usually classified as proper nouns'

b. I per arribar a aquesta hipòtesi de treball fa un *esforç de síntesi* que descansa en el diàleg, la interrelació, les concomitàncies,

(CTILC, 2004)

'And to reach this working hypothesis, he makes a synthesis effort which lies on the dialogue, interrelations and concomitances'

c. les màquines faran *la feina dels pagesos* i ningú voldrà treballar als conreus.   (CTILC, 2003)

182

'the machines will do the peasant's job and nobody will work in farming'

d. Pel camí vaig fent *les preguntes de rigor*.   (CTILC, 2010)

'I make the obligatory questions on my way'

e. Delfí li va dir que ell i Anna estaven fent *una visita d'inspecció* per vigilar que tot s'ordenés de la manera correcta.   (CTILC, 2010)

'Delfí told him that he and Anna were doing an inspection visit to check that all was correctly organised'

f. quan van fer *el viatge de noces*.   (CTILC, 2008)

'when they went on their honeymoon trip'

g. També cal obligar-los que facin *una anàlisi de les qüestions de fons* que han provocat aquest distanciament de la societat balear.

(CTILC, 2003)

'They must be also obliged to make an analysis of the different fundamental matters which have caused this rift in the Balearic society'

h. Mendelsohn va fer *uns dibuixos de caràcter «tel·lúric i planetari»* i els investigadors varen donar-li el vistiplau.   (CTILC, 2006)

'Mendelsohn make some telluric and planetarian drawings and the researchers gave him their approval'

i. Quan, al cap de pocs anys, mor Bohr, algú fa *una fotografia de la pissarra del seu estudi*:   (CTILC, 2016)

'When after a few years Bohr dies, someone takes a picture of his study's blakboard'

j. Amb aquesta metàfora ens podríem trobar que una cultura *fa l'abraçada de l'ós* a una altra cultura, de forma que l'ofega i l'anul·la. (CTILC, 2005)

'With this metaphor we could find a culture that give the bear hug to another culture, in a way that it suffocates and cancels the other'

(136)   L'Eloi sempre li fa *preguntes que la deixen pensant uns quants dies*.

(CTILC, 2012)

'Eloi always asks questions that leave her thinking for some days'

As in the case of *take*-LVCs, the different kinds of PP modifiers are relevant for the analysis. Some Romance PP modifiers can be assimilated to the genitive in Germanic languages (Bonet & Solà, 1986), as in *la feina dels pagesos* (135c), *el viatge de noces* (135f) or *fa l'abraçada de l'ós* (135j); while others can be considered quasi-arguments

holding different thematic roles (as proposed in Levin & Rappaport Hovav, 2005), such as Theme in *una fotografia de la pissarra del seu estudi* (135i). Also, some instances are attributive PPs introducing qualities and description of the noun, as in *uns dibuixos de caràcter «tel·lúric i planetari»* (135h).

A closer look to the LVC with higher number of instances (more than a 100 in total) shows only one instance of high co-occurrence with a specific adjective. This is the case of *fer esforç* with the adjective *gran* ('big', in 25/79, 31.65%), as in (137).

(137)   havia de fer un *gran* esforç per no abraçar-te contra el meu cor.

(CTILC, 2001)

'I had to make a great effort not to hug you against my heart'

In a similar note, *fer un viatge* co-occurs mainly with the adjective *llarg* ('long', in 11/45, 24.44%), as in (138).

(138)   quan fem un viatge *llarg* a fusos horaris molt diferents del nostre.

(CTILC, 2001)

'when we go on a long trip to very different time zones to ours'

In other two cases, there are two different adjectives co-occurring with the same LVCs, as it is the case in *fer ús* with the opposite adjectives *mal* ('bad', in 7/45, 15.56%) and *bon* ('good', in 6/45, 13.33%), as in (139); as well as *fer anàlisi* with more specific adjectives such as *exhaustiva* ('thorough', in 5/59, 8.47%) and *rigorosa* ('rigurous', in 5/59, 8.47%), as in (140).

(139)   a. Sempre és millor conèixer que no conèixer, malgrat el fet que es pugui fer un *mal* ús d'aquest coneixement.   (CTILC, 2007)
'It is always better to know than not to know, although the fact that one can make a bad use of this knowledge'
b. no significa que siguem capaços de fer un *bon* ús d'aquestes habilitats ni que siguem prou humans.   (CTILC, 2004)
'that does not mean that we are able to make a good use of this capacities nor that we are human enough'

(140)   a. Els professors de Cambridge pensaven que el primer objectiu de la filosofia és fer una anàlisi *exhaustiva* del llenguatge. (CTILC, 2003)

'The Cambridge Professors though that the philosophy's first objective is to make an exhaustive analysis of language'

b. Fer una anàlisi *rigorosa* del concepte de poder demana entendre les variables que el componen. (CTILC, 2015)

'To make a rigorous analysis of the concept of power involves understanding the variables that conform it'

Finally, in *fer pregunta* there is no adjective appearing in more than two instances. Thus, the range of adjectives combined with this LVC is varied and no co-occurrence is to be found.

In the case of *fer*-LVCs, there is an even distribution between LVCs accepting bare nominals and LVCs with obligatory determiners (see §3.4.3.1). In both cases, relational and qualitative adjectives are found.

3.6.2.5. Spanish hacer-LVCs

Figure 45 shows that modification in *hacer*-LVCs is varied and lemma-dependent. On the one hand, there are some lemmas with a strong predominance of modification, as in *hacer énfasis* (91.96%), *hacer uso* (90.91%), or a clear preference, in *hacer análisis* (69.85%) and *hacer colección* (57.81%). On the other hand, the rest of LVCs have less than half of their instances with modification: *hacer esfuerzo* (48%), *hacer crítica* (41.33%), *hacer carrera* (39.67%), *hacer ejercicio* (37%), *hacer declaración* (32.32%), *hacer pregunta* (27%), *hacer broma* (26.67%), *hacer viaje* (23.35%), and even less often in *hacer referencia* (9.50%).
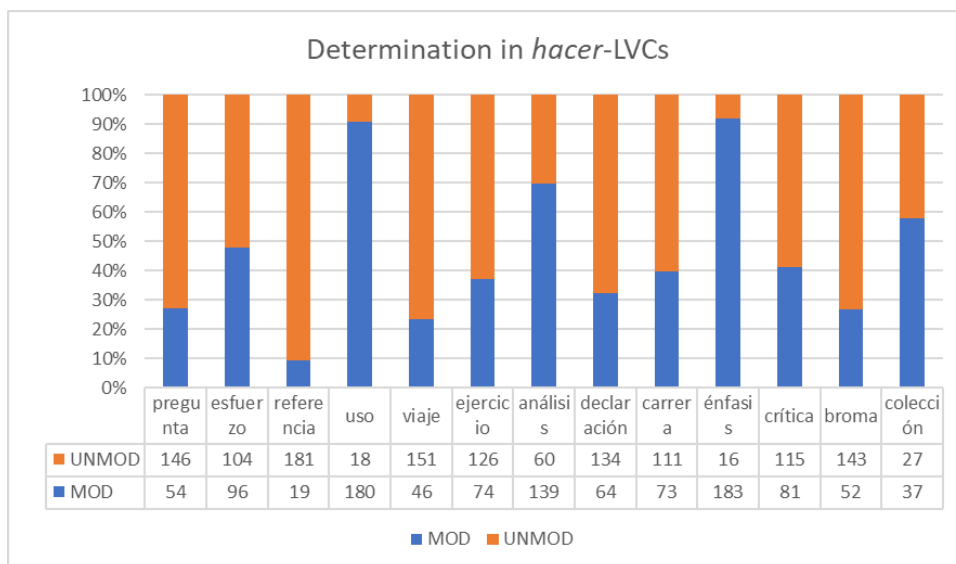
Figure 45. Distribution of modification in Spanish hacer-*LVCs per lemma*

Table 48 shows that adjectival modification is only predominant in those lemmas where modified instances are not the majority.

Modification	pregunta	esfuerzo	referencia	uso	viaje	ejercicio	análisis	declaración	carrera	énfasis	crítica	broma	colección
Unmod (No)	146 (73%)	104 (52%)	181 (90.50%)	18 (9.09%)	151 (76.65%)	126 (63%)	60 (30.15%)	134 (67.68%)	111 (60.33%)	16 (8.04%)	115 (58.67%)	143 (73.33%)	27 (42.19%)
Adj	43 (21.50%)	80 (40%)	14 (7%)	8 (4.04%)	32 (16.24%)	34 (17%)	45 (22.61%)	37 (18,69%)	52 (28.26%)	2 (1.01%)	49 (25%)	18 (9.23%)	15 (23.44%)
PP	5 (2.5%)	11 (5.5%)	5 (2.5%)	145 (73.23%)	11 (5.58%)	33 (16.50%)	69 (34.67%)	22 (11.11%)	17 (9.23%)	162 (81.40%)	27 (13.78%)	31 (15.90%)	17 (26.56%)
Rel	4 (2%)	-	-	1 (0.51%)	1 (0.51%)	3 (1.5%)	2 (1.01%)	-	2 (1.09%)	-	-	2 (1.03%)	1 (1.56%)
NC	-	-	-	-	-	-	-	-	-	-	-	-	-
Gen	-	-	-	-	-	-	-	-	-	-	-	-	-
Various	2 (1%)	5 (2.5%)	-	26 (13.13%)	2 (1.02%)	4 (2%)	23 (11.56%)	5 (2.52%)	2 (1.09%)	19 (9.55%)	5 (2.55%)	1 (0.51%)	4 (6.25%)
TOTAL	200	200	200	198	197	200	199	198	184	199	196	195	64

Table 48. Distribution of modification subclasses in *hacer*-LVCs per lemma

Instead, the most often modified LVCs present higher proportions of PP modification than adjectival modification, especially in *hacer uso* (73.23% of PP

modification versus 4.04% of adjectival modification), and *hacer énfasis* (81.04% of PP modification versus 1.01% of adjectival modification), as in (141). The preposition selected for both nouns is, however, divergent: *hacer uso* selects *de* 'of' and *hacer énfasis* selects *en* 'in'.

(141)   a. Las generalistas suelen hacer *uso de una amplia variedad de hábitats*,

                                                        (CORPES XXI, 2001)

'The generalists usually make us of a wide variety of habitats'

b. la empresa hizo *énfasis en su tecnología de clustering* con el eslogan "Irrompible".   (CORPES XXI, 2001)

'the company made emphasis in its clustering technology with the slogan "Unbreakable"'

Also, other LVCs have a relevant presence of PP modification, as *hacer análisis* (34.67% of PP modification versus 22.61% of adjectival modification), and *hacer colección* (26.56% of PP modification versus 23.44% of adjectival modification), as in (142). The nature of these PPs will be of interest in the upcoming discussion (§3.6.3).

(142)   a. que recorre todo el país haciendo *análisis de las carreteras* para garantizar la calidad de las estructuras   (CORPES XXI, 2003)

'that goes all around the country making analysis of the roads to confirm the structures' quality'

b. Siempre quisiste hacer *una colección de sombreritos bolivianos*.

                                                        (CORPES XXI, 2002)

'You always wanted to make a Bolivian hat collection'

In other lemmas, PP modification is also relevant, but to a lesser extent: 5.5% in *hacer esfuerzo*, 5.58% in *hacer viaje*, 16.50% in *hacer ejercicio*, 9.23% in *hacer carrera*, and 13.78% in *hacer crítica,* as in (143).

(143)   a. Para asistir hoy a una boda o entierro religiosos, hay que hacer *un esfuerzo de abstracción* de lo que allí se esta haciendo,

                                                        (CORPES XXI, 2001)

'To attend a religious wedding or funeral today, one must make an abstraction effort of what is happening there'

b. Durante los tres meses siguientes, cada vez que Juan hacía *un viaje de fin de semana* a Bilbao (unas dos veces al mes), (CORPES XXI, 2001)

'During the next three months, every time Juan made a weekend trip to Bilbao (one or twice a month)'

c. Más nervioso que él se puso Sabrina, que hizo *un ejercicio de contención* para no derrumbarse en lágrimas.   (CORPES XXI, 2001)

'More nervous than him was Sabrina, who make a contention exercise not to break down in tears'

d. ¿Que haría *una carrera de camellos* en el desierto, frente a las pirámides?   (CORPES XXI, 2009)

'Would you do a camels' run in the desert in front of the pyramids?'

e. se hacían *las críticas del trabajo*, se veía cuáles habían sido los avances y planteábamos nuevas directrices para varios meses.

(CORPES XXI, 2003)

'job critics were made, which showed what were the progresses and we outlined new guidelines for some months'

As in the case of Catalan *fer*, the PP modification has a special relevance in *hacer*-LVCs, and PP modifiers of different nature have been identified: genitive PPs, as in *una carrera de camellos* 'a camel race' (143d); quasi-argumental PPs, as in *un esfuerzo de abstracción* 'an effort of abstraction' (143a) or *un ejercicio de contención* 'an exercise of contention' (143c); as well as attributive PPs, as in *un viaje de fin de semana* 'a weekend trip' (143b).

An analysis of the co-occurrence of Spanish *hacer*-LVCs with specific adjectives shows that there are certain tendencies which might have an effect in the percentages of modification. This is the case of three LVCs (*hacer referencia* 'make reference', *hacer uso* 'make use' and *hacer énfasis* 'make emphasis') which present low percentages of adjectival modification, but they combine in most cases with specific adjectives: *hacer referencia* (144) is mostly combined with *directa* ('direct', 3/14, 21.43%), *hacer uso* (145) with *mejor* (5/34, 14.71%), and especially *hacer énfasis* (146) with *especial* ('special', 13/21, 61.90%).

188

(144)   hacen referencia *directa* a las imágenes, y señalo especialmente a los pies

de foto.   (CORPES XXI, 2001)

'make direct reference to image, and I point specially towards image

captions'

(145)   y la necesidad de hacer un *mejor* uso de los recursos en general en la

educación   (CORPES XXI, 2002)

'and the need to make a better use of the resources in Education in

general'

(146)   Añadió en que se hará *especial* énfasis en que los canes menores de tres

meses de edad puedan ser vacunados en esa campaña

(CORPES XXI, 2001)

'He added that special emphasis will be made in that the younger than 3

months hounds could be vaccinated in this campaign'

In contrast, the LVCs with higher percentages of adjectival modification (*hacer esfuerzo* 'make (an) effort'*, hacer un análisis* 'make an analysis' and *hacer carrera* 'make a career'*) show a wider range of combinatory patterns, although some coappearances are more habitual. For instance, *hacer esfuerzo* combines with two adjectives repeatedly, as in (147): *gran* ('big', 28/85, 32.94%), and *enorme* ('huge', 10/85, 11.76%).

(147)   a. Hice un *gran* esfuerzo para no descomponer el gesto.

(CORPES XXI, 2001)

'I made a big effort not to break the gesture down'

b. de que realmente se puede hacer un esfuerzo *mayor* en ahorrar en

recursos mal utilizados, evitando así alzas impositivas.

(CORPES XXI, 2002)

'that a big effort can be really made in saving in badly used resources, thus

avoiding tax rises'

c. la selección deberá hacer un esfuerzo *enorme*, aprovechar la fuerza física

de nuestros jugadores,   (CORPES XXI, 2002)

'the national team could make a huge effort to make the most of our

players' physical strength'

In the case of *hacer carrera*, the most repeated adjectives are *política* ('political', 7/54, 12.96%) and *universitaria* ('academic', 6/54, 11.11%), as in (148).

(148)   a. ¿O vos te creés que mi carrera *política* la hice a base de mi labia?

(CORPES XXI, 2002)

'Or do you believe that I made my political career for my loquacity?'

b. en la imposibilidad de hacer la carrera *universitaria*.

(CORPES XXI, 2009)

'in the impossibility to make a career in the university'

Finally, *hacer análisis* combines repeatedly only with *detallado* ('detailled', 5/68, 7.35%), as in (149).

(149)   El primer bloque se dedica a hacer un análisis *detallado* del
funcionamiento de la industria   (CORPES XXI, 2004)
'The first section focuses on making a detailed analysis of the industry
functioning'

In contrast, *hacer colección* ('make (a) collection') is a low-frequency LVC with more frequent modification than the rest but it does not show any tendency of co-occurrence with specific adjectives. Some adjectives appear only twice at most: *buena* ('good'), *primera* ('first') and *gran(de)* ('good'), as illustrated in (150).

(150)   a. Cuando los diseñadores hacemos una *buena* colección,

(CORPES XXI, 2010)

'When us designers make a good collection'

b. a usted también la convenció la ex directora de Vogue, Diana Vreeland,
para hacer su *primera* colección.   (CORPES XXI, 2007)

'you were also convinced by the former Vogue's director, Diana Vreeland,
to make your first collection'

c. Nosotros hacíamos *grandes* colecciones a beneficio de los ancianos de
Conapran en el Palacio Cousiño.   (CORPES XXI, 2014)

'We made great collections benefiting Conapran's elders in Palacio
Cousiño'

Also, the fact that most *hacer*-LVCs accept bare nominals as their NVE does not seem to influence the kind of modification found, since both relational and qualitative adjectives are present in all LVCs, regardless of their adjectival modification rates. Relational adjectives are further found in LVCs with an optional determiner (151), as well as with an obligatory determiner (152).

(151) a. no se hará uso *legal* de ella si no es estrictamente necesario.

(CORPES XXI, 2002)

'no legal use of it will be made unless it is strictly necessary'

b. Cuando escribo no puedo parar todo y dedicarme a hacer crítica *literaria* durante un mes.   (CORPES XXI, 2001)

'When I write I cannot stop everything and focus on doing literary criticism for a month'

(152) a. y hago el único esfuerzo *intelectual* que me permite mi estado a esas horas: leerme el periódico por encima.   (CORPES XXI, 2001)
'and I make the only intellectual effort that allows my condition at this time: skimming the newspaper'
b. Y se hizo una pregunta *retórica*   (CORPES XXI, 2002)
'And a rhetorical question was made'
c. querría hacer una declaración *institucional* para reiterar el compromiso de España de solidaridad con Argentina   (CORPES XXI, 2001)
'I would love to make an institutional declaration to confirm the solidarity agreement in Spain with Argentina'

In sum, the modification in *tomar*-LVCs shows similar tendencies to those of *fer*-LVCs in Catalan, showing an alignment of these two languages regarding the behaviour of this verb.

3.6.3. Discussion

The analysis of determination in *make*-LVCs presents all four languages with more than 50% of NVEs introduced by a determiner. As in the other two LVs, English *make*-LVCs presents the highest percentage of determined NVEs (82.58%). The rest of languages are ordered as follows: Catalan *fer*-LVCs (69.76%), German *machen*-LVCs (57.31%), and Spanish *hacer*-LVCs (53.62%). In English, only one LVC frequently

combines with bare singulars, *make progress*, while German *machen*-LVCs present some signs of lexicalization with bare plurals (i.e. *Sorgen machen, Gedanken machen* and *Angaben machen*). In the two Romance languages, the more frequent LVCs appear more often followed by a bare nominal (*fer referència / hacer referencia* 'make reference'*, fer ús/ hacer uso* 'make use', and also *fer feina* 'do work' in Catalan and *hacer ejercicio* 'do exercise' in Spanish).

When the correlation between determination and modification is examined, the same tendency is found as in the previous two LVs: bare nominals are more often modified in English, although this is the language with fewer instances of bare NVEs, while the other three languages have the lowest proportion of bare nominals with modification.

Regarding modification, English *make*-LVCs shows the highest proportion (60.85%) followed by the two Romance languages (45.19% in Spanish *hacer*-LVCs and 37.25% in Catalan *fer*-LVCs). German *machen*-LVCs are the least modified (28.36%). All languages present an increasing tendency that goes from high-frequency LVCs to low-frequency LVCs, with the latter being the most often modified.

For this LV, the two Romance languages display the coexistence of two subclasses of modification: PP modification and adjectival modification, while the predominant subclass in the two Germanic languages is adjectival modification, as with the other two LVs. Unlike the two previous LVs, adjectival modification cannot be compared with previous studies in any language, because there are no corpus-based studies focusing on the LV *make*.

In English, the most modified *make*-LVCs are the most frequent, but there are no co-occurrences with specific adjectives which could influence the overall results because of their tendency to co-occur. Besides, there is one lemma *make mention* which selects PP modification (with *of*) over adjectival modification. Likewise, modification in German *machen*-LVCs is mainly found in the low-frequency lemmas, while the high-frequent LVCs appear mostly unmodified and they show some signs of lexicalization, since they mainly select bare plurals: *Gedanken machen* ('make thoughts, lit. to reflect')*,* and *Sorgen machen* ('make worries, lit. to worry')*,* contrary to *Angaben machen* ('make details, to give information') which tends to co-occur with two adjectives: *falsch* ('false') and *detalliert* ('detailed').

In the two Romance languages, some co-occurrences are found, especially in cases with fewer instances such as the Spanish *hacer referencia* ('make reference') with the adjective *directa* ('direct'), *hacer uso* ('make use') with *mejor* ('better'), and *hacer énfasis* ('make emphasis') with *especial* ('special'). In both languages, however, the presence of PP modification is relevant in several lemmas: *fer ús* ('make use') and *fer un viatge* ('make a trip') in Catalan; *hacer uso* ('make use'), *hacer análisis* ('make analysis'), *hacer énfasis* ('make emphasis'), *hacer broma* ('make a joke') and *hacer colección* ('make a collection'). The kind of PPs selected in each lemma is of different nature, but most of them are genitive modifiers and the so-called thematic quasi-arguments.

As discussed in previous verbs (specially in *take,* §3.5.3), the kinds of PP are relevant for the analysis, as well as the discussion of its relation to the NVE or the LVC as a whole. Some of the Romance PP modifiers are equivalent to the genitive in Germanic languages (as pointed out in Bonet & Solà, 1986, for Catalan). That is the case for some instances in Catalan (i.e. *la feina dels pagesos* 'the peasants' job', *el viatge de noces* 'the honeymoon', and *l'abraçada de l'ós* 'the bear's hug') and in Spanish (i.e. *una carrera de camellos* 'a camels' race').

In contrast, other PP modifiers are considered quasi-arguments (as proposed in Levin & Rappaport Hovav, 2005) with different thematic roles. For Cremades (2016, 2017), nouns express events and allow argument selection – more specifically, the realization of direct internal arguments. She argues that these PPs are generated in a sister position of the unspecified lexical root of the noun, and their origin is that of the DO of the verbal base in deverbal nouns. Therefore, it is related to the transitivity of the verbal base.

In the present corpus sample, some of the selected LVCs for the verb *make* (and its corresponding translation in the other three languages) show a behaviour that confirms Cremades' proposal for the quasi-arguments in Catalan eventive nouns. This is the case for some Catalan *fer*-LVCs (i.e. *fer una fotografia de la pissarra* 'take a picture of the blackboard') and Spanish *hacer*-LVCs (i.e. *hacer uso de una amplia variedad de hábitats* 'make use of a wide variety of habitats'), but *also* in English *make*-LVCs (i.e. *makes no*

mention of the United States) and German (i.e. *Angaben über Lage* 'details about the situation')[30].

Other subclasses of modification point towards the relevance of the verbal base in the NVEs for the selection of so-called quasi-arguments. It is the case of relative clause modification found in several nouns in English *make*-LVCs and German *machen*-LVCs. More specifically, some of these relative clauses seem to have an origin within the argument selection of the verbal base, as in *make the assumption that* and *make the claim that*; the clause introduced by *that* corresponds to the direct object of the base verbs *to assume* and *to claim*.

The selection of NVEs with a verbal transitive base in *make*-LVCs has proven to be a relevant variable for the modification possibilities in all languages. The influence of the verb base in the selection of modifiers or quasi-arguments will be relevant for the proposed typology in Chapter 6.

Moreover, noun classifiers have proven to be more present in the English *make*-LVCs than in the two previous LVs (*give* and *take*) due to their presence in the modification of certain nouns: *call, claim* and *payment.* More specifically, *make a call* is the LVC with a higher presence of such noun classifiers with its co-occurrence with *phone* (in *phone call*); in this case, there is a clear sign of lexicalization, as it is present in 37.37% of *make a call* instances. Although noun classifiers are considered as part of LVCs modification in this dissertation, the analysis of noun classifiers in Distributed Morphology is that of an incorporation of the modifier with the root of the head of the compound (since Harley, 2009). With this in mind, the co-occurrence of *phone* and *call* (as well as *asylum* and *claim*, or *mortgage* and *payment*) should not be treated as a sign of lexicalization of the construction but rather of regular primary compounds typical in the English language.

In sum, although some language-specific and lemma-specific properties and tendencies are identified, adjectival modification is still the most predominant in all languages and the variety of adjectives selected is wide. However, the rest of modification

---

[30] As the example in German shows, not all PP modifiers are introduced by the preposition *of/de/von*: other prepositions can also be found and they are also dependent on the verbal base (as pointed out in Cremades, 2016: 120).

subclasses are especially relevant in this LV because they provide proof of the influence of the verbal base in the modification selection of the NVE in LVCs.

3.7. Statistical analysis and discussion

In order to confirm the corpus-based, empirical trends attested in the previous sections, the statistical method employed is the *binary regression analysis*, which is a statistical technique that uses explanatory variables (which can be categorical or scalar) to estimate their effect on the linguistic variable (which must be categorical). That is, this statistical analysis aims to determine the "chances of a construction A to be chosen in a particular type of context compared with the chances of a construction B to be used in the same type of context" (Levshina, 2015: 253). The aim of the implementation of this analysis is to observe the relation between the response variable (in this case, the modification) and the other variables examined (which are called predictor variables). With this methodological analysis, relevant predicting factors will be identified and discussed.

Binary logistic regression is chosen for the analysis of the object of study of this corpus-based research since the response variable (modification) is not numerical and has only two possible values (that is, yes/no or modified/unmodified). The purpose to implement this statistical test is to examine the extent to which the morphosyntactic variables considered in the previous sections (number and determination) as well as the most general variables (LV and frequency) predict or explain the patterns of modification attested.

In logistic regression, we look at both the general performance of a model as well as at individual coefficients showing the effect of the predictor variables on the outcome of interest. 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"[31].

Since the whole corpus data is too extensive, the analysis has been performed on each language separately. This way the overall statistics allow us to see the effects that individual predictors have in each language.

---

[31] Available at: http://corpora.lancs.ac.uk/stats/toolbox.php

Before starting the discussion on the output of the model, the most important measures should be commented. First, the Model Likelihood Ratio shows the significance with a p-value (with a 95% confidence level, as standard), which tests whether the deviance of the model is significantly different from that of a model without predictors (Levshina, 2015: 258). The other two relevant measures (C-index and Nagelkerke R^2) measure the accuracy and predictive power of the model (Speelman, 2014). The C-index values range from 0 to 1, with a threshold of 0.7 to have an excellent discriminatory power (Brezina, 2018).

For the English dataset, Table 49 shows that the logistic regression model was statistically significant, Likelihood ratio test (LL): 160.07 (p < 0.0001): although the model accuracy should be taken cautiously as the C-index presents low accuracy results (0.58). The model explained 3% (Nagelkerke R^2) of the variance in modification. The outcome variable is modification, the first thing to bear in mind is that the reference level of the response variable is the first one: unmodified (a_unm). Therefore, the coefficients of the predictor variables (LV, Det, Num, Freq) have to be interpreted in relation to the second level of the response variable, i.e. modified (b_mod).

Logistic Regression Model

<u>Overall model statistics:</u>

Overall significance: Likelihood ratio test (LL): 160.07 (p < 0.0001);

Model accurary: C-index: 0.58;

Predictive power: Nagelkerke R^2: 0.03;

AIC: 9598.13

<u>Coefficients:</u>

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	0.212	0.056	3.787	0.000	1.236	1.108	1.380
LVb_take	0.503	0.063	8.028	0.000	1.653	1.463	1.869
LVc_make	0.409	0.061	6.717	0.000	1.506	1.336	1.697
Detb_no	0.358	0.066	5.392	0.000	1.431	1.257	1.631
Numb_plur	-0.425	0.060	-7.069	0.000	0.654	0.581	0.735
Freqb_medium	0.051	0.058	0.894	0.371	1.053	0.941	1.179
Freqc_low	-0.355	0.062	-5.765	0.000	0.701	0.621	0.791

Table 49. Output of the logistic regression model for the English dataset

Let us now briefly comment on the significance of the variables and the interpretation of their coefficients. First, the p-value will be taken into account when commenting on the relevant variables (≤0.05), and it will be combined with the estimate (log odds) which expresses the probability of the specific variable to be interchangeable.

The variable LV (*take* and *make*) show a positive estimate (log odds), which indicates that they increase the probability of modification compared to the reference level of this variable, i.e. the LV *give*. More specifically, *take*-LVCs are 1.653 more likely (estimate, odds) than *give*-LVCs to be modified. Likewise, *make*-LVCs are 1.506 more likely than *give*-LVCs to be modified.

Also, the variable Det=b_no (bare nouns) shows a positive estimate (log odds), which indicates that it significantly inncreases the probability of modification compared to the reference level of this variable, i.e. Det=a_yes (which are NVEs introduced by a determiner). More specifically, it is 1.431 more likely (estimate, odds) than determined LVCs to be modified. In this case, however, it is important to take into account that this variable includes bare nominals in both singular and plural number.

Finally, plural NVEs (Num=b_plur) show a negative estimate and, thus, are less likely than singulars to be modified, which is in this case 0.654 lower.

Regarding the predictor variable frequency, it also shows relevant results, but to a lesser extent, as it is only statistically significant in the case of low-frequency LVCs. In this case, Freq=low shows a negative estimate which indicates that it is less likely to be modified than the reference variable (Freq=high), more specifically 0.701 less likely.

This statistical analysis is relevant because all variables have been compared in relation to the response variable (that is, modification). These statistical results will be compared to those of the rest of the languages under study, in order to compare which variables are relevant for modification in each dataset.

For the German dataset, Table 50 shows that the logistic regression model was statistically significant, Likelihood ratio test (LL): 409.73 ($p < 0.0001$), and the model accuracy is acceptable, as the C-index is 0.73. The model explained 17% (Nagelkerke R^2) of the variance in modification. As in the previous analysis, the outcome variable is modification, and the response variable is Mod=b_mod (that is, modified NVE).

Logistic Regression Model

Overall model statistics:

Overall significance: Likelihood ratio test (LL): 409.73 (p < 0.0001);

Model accurary: C-index: 0.73;

Predictive power: Nagelkerke R^2: 0.17;

AIC: 3270.57

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	-1.487	0.109	-13.594	0	0.226	0.182	0.279
LVb_take	-1.924	0.173	-11.091	0	0.146	0.103	0.204
LVc_make	-0.375	0.102	-3.675	0	0.687	0.562	0.839
Detb_no	0.393	0.109	3.599	0	1.482	1.197	1.837
Numb_plur	0.533	0.129	4.147	0	1.705	1.325	2.194
Freqb_medium	0.606	0.119	5.094	0	1.833	1.452	2.315
Freqc_low	1.006	0.115	8.763	0	2.735	2.185	3.428

Table 50. Output of the logistic regression model for the German dataset

The variable LV in the German dataset (*nehmen* and *machen*) shows a negative estimate (log odds) which indicates that they decrease the probability of modification compared to the reference level of this variable, i.e. the LV *give/geben*. More specifically, *nehmen*-LVCs are 0.146 less likely (estimate, odds) than *geben*-LVCs to be modified. Likewise, *machen*-LVCs are 0.687 less likely than *geben*-LVCs to be modified.

In parallel, the variable Det=b_no (bare nouns) shows a positive estimate (log odds), which indicates that it significantly increases the odds of modification compared to the reference level of this variable, i.e. Det=a_yes (which are NVEs introduced by a

determiner). More specifically, bare nouns are 1.482 more likely (estimate, odds) than determined LVCs to be modified. In this case, it also includes both bare singulars and bare plurals.

Also, plural NVEs (Num=b_plur) show a positive estimate and, thus, are more likely than singulars to be modified, in this case 1.705 more likely.

Finally, frequencies also show relevant results: Freq=medium is 1.833 more likely to be modified than the reference variable (i.e. high frequency), and Freq=low is 2.735 more likely than high frequency German LVCs.

As already observed in previous sections in this chapter, English and German LVCs present similar as well as divergent tendencies with regard to modification. In the statistical relevance of the variables involved, bare nouns are more likely to be modified in both languages, while the rest of the variables present divergent tendencies.

Let us now discuss the statistical regression analysis results for the two Romance languages. For the Catalan dataset, Table 51 shows that the logistic regression model was statistically significant, Likelihood ratio test (LL): 104.82 (p < 0.0001); although the model accuracy should be taken cautiously as the C-index presents results just below the threshold (0.61). The model explained 5% (Nagelkerke R^2) of the variance in modification. Again, the outcome variable is modification, and the predictor variables (LV, Det, Num, Freq) will be interpreted in relation to the response variable.

Logistic Regression Model

Overall model statistics:

Overall significance: Likelihood ratio test (LL): 104.82 (p < 0.0001);

Model accurary: C-index: 0.61;

Predictive power: Nagelkerke R^2: 0.05;

AIC: 3707.78

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	-0.710	0.089	-7.959	0.000	0.491	0.412	0.585
LVb_take	0.184	0.126	1.463	0.143	1.202	0.939	1.536
LVc_make	0.305	0.096	3.173	0.002	1.356	1.124	1.639
Detb_no	-0.637	0.090	-7.091	0.000	0.529	0.443	0.630
Numb_plur	0.114	0.094	1.207	0.227	1.120	0.931	1.347
Freqb_medium	-0.079	0.095	-0.828	0.408	0.924	0.767	1.113
Freqc_low	0.412	0.113	3.637	0.000	1.509	1.208	1.883

Table 51. Output of the logistic regression model for the Catalan dataset

In this case, the only statistically significant variables are the LV *fer* ('make'), Det=no (bare nouns) and Freq=low, while the variable Num (sing/plur) does not show significant results in this logistic regression model for the Catalan dataset.

The variable LV in the Catalan dataset only shows a significant p-value in the case of *fer* which shows a positive estimate (log odds): this LV increases the odds of modification compared to the reference level of this variable, i.e. the LV *donar* ('donar'). More specifically, *fer*-LVCs are 1.356 more likely (estimate, odds) than *donar*-LVCs to be modified.

In contrast, the variable Det=b_no (bare nouns) shows a negative estimate (log odds), which indicates that it significantly decreases the probability of modification compared to the reference level of this variable, i.e. Det=a_yes (which are NVEs introduced by a determiner). More specifically, it is 0.529 less likely (estimate, odds) than determined LVCs to be modified. In this case, it also includes both bare singulars and bare plurals.

Finally, frequencies show significant results only in the low frequency group. In this case, Freq=low is 1.509 more likely to be modified than the reference variable (i.e. high frequency). The results for the medium-frequency group do not reach statistical significance in comparison to the reference variable (p= 0.408).

For the Spanish dataset, Table 52 shows that the logistic regression model was statistically significant, Likelihood ratio test (LL): 290.88 (p < 0.0001); although the model accuracy should be taken cautiously as the C-index is below the threshold (0.62). The model explained 5% (Nagelkerke R^2) of the variance in modification. Again, the outcome variable is modification, and the predictor variables (LV, Det, Num, Freq) will be interpreted in relation to the response variable.

Logistic Regression Model

Overall model statistics:

Overall significance: Likelihood ratio test (LL): 290.88 (p < 0.0001);

Model accurary: C-index: 0.62;

Predictive power: Nagelkerke R^2: 0.05;

AIC: 9522.21

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	-0.792	0.062	-12.847	0.000	0.453	0.401	0.511
LVb_take	0.171	0.062	2.773	0.006	1.187	1.051	1.340
LVc_make	0.813	0.059	13.706	0.000	2.256	2.008	2.535
Detb_no	-0.341	0.054	-6.318	0.000	0.711	0.640	0.790
Numb_plur	-0.280	0.057	-4.895	0.000	0.756	0.676	0.845
Freqb_medium	0.010	0.065	0.147	0.883	1.010	0.889	1.146
Freqc_low	0.068	0.060	1.138	0.255	1.070	0.952	1.203

Table 52. Output of the logistic regression model for the Spanish dataset

In this case, the only statistically significant variables are the LVs *tomar* ('take') and *hacer* ('make'), Det=no (bare nouns) and Num=plur; however, the variable frequency does not show significant results in this logistic regression model for the Spanish dataset.

First, the variable LV in the Spanish dataset shows significant p-values and positive estimates for both LVs: *tomar* and *hacer* increase the probability of modification in comparison to the reference level for this variable (i.e. the LV *dar*, 'give'). More specifically, *tomar*-LVCs are 1.187 more likely to be modified than *dar*-LVCs, and *hacer*-LVCs are 2.256 more likely.

In contrast, the variable Det=b_no (bare nouns) shows a negative estimate (log odds), which indicates that it significantly decreases the probability of modification compared to the reference level of this variable, i.e. Det=a_yes (which are NVEs introduced by a determiner). More specifically, it is 0.711 less likely (estimate, odds) than determined LVCs to be modified. In this case, it also includes both bare singulars and bare plurals.

Also, plural NVEs (Num=b_plur) show a negative estimate and, thus, are less likely than singulars to be modified, which is in this case 0.756 higher.

Finally, neither medium nor low frequencies dshow significant results in comparison to the reference variable (i.e. high frequency).

In sum, when all statistical results are taken into account, the analysis of the variable modification has proven to be significant in all four languages under study. However, not all models are highly accurate; in this case, only the German dataset has had a statistically relevant accuracy, while the models for English, Catalan and Spanish datasets present a low accuracy (with C-index lower than 0.7). The outcome variable (modification) and the predictor variables are the same throughout all the regression models, i.e. light verbs (LV), determination (Det), number (Num) and frequency (Freq).

First, when the LV variable is observed, the LVs *take* and *make* (and their corresponding verbs in each language) are more likely than *give* to be modified in English, Catalan and Spanish. For German, however, the odds are reversed: these verbs are less likely to be modified. This variable singles the German LVCs database out, as the quantitative analysis in previous sections (§3.4, §3.5 and §3.6) already pointed out: *geben*-LVCs are modified in 31.76% of instances, while *nehmen*-LVCs in 5.71% and *machen*-LVCs in 28.36%.

Second, the determination variable presents opposite trends in the two language families: Germanic and Romance languages. On the one hand, English and German point towards bare nouns (both singular and plurals) to be more likely modified than NVEs introduced by a determiner. On the other hand, Catalan and Spanish have reversed results, as bare nouns in these Romance languages are less likely modified than determined nouns. These results are in line with the general properties of the nominal domain in these language families, which will be further discussed in Chapters 5 and 6.

Third, number of the NVE presents heterogeneous tendencies. English and Spanish datasets show that plural number in the NVE are less likely than singular NVEs to be modified. However, the German dataset shows the opposite trend, as plurals are more likely to be modified. As for the Catalan dataset, it does not provide significant statistical trends.

Finally, the variable frequency is also divergent, but the regression model only suggests significant trends in the low frequency group in comparison to the high frequency (taken as the reference variable). Hence, the low frequency LVCs are less likely to be modified in English, while they are more likely modified in German and Catalan. The probability of modification for the Spanish dataset in relation to frequency is not significant.

To conclude the statistical models' analysis, these results highlight and confirm those trends already identified in the analysis of individual LVs in the previous sections (cf. §3.4, §3.5 and §3.6).

3.8. Conclusion and final remarks

The final section of this chapter will first briefly summarize the findings of the corpus-based study and later outline the main conclusions with some final remarks for theoretical linguistics.

The summary of the quantitative findings will focus on both the similarities and divergences found between the three verbs and among the four languages under study. The focus of the analysis has been put on the weight of the NVE, more specifically on its determination and modification.

First, the results on determination have shown similar trends in the four languages in two verbs: *give*-LVCs and *make*-LVCs have a majority of NVE introduced by a determiner in all languages, while *take*-LVCs have predominantly determined NVEs in English and Spanish but not in German and Catalan with more bare nouns than determined NVEs with the LVs Ger. *nehmen* and Cat. *prendre* ('take'). The differences in determination are, thus, not in relation to language families.

When bare nouns are taken into consideration, however, there are general tendencies in the corpus database which can be related to differences in language families. That is, the two Germanic languages present more often modified bare nouns as NVEs,

while the two Romance languages have a more restricted modification in bare NVEs. This trend is supported by the logistic regression model presented (§3.7).

Second, the results of modification have shown some divergences between the four languages analysed. English LVCs has the highest proportion of modified NVEs, while German has the lowest – especially the *nehmen*-LVCs, which have shown residual proportion of modification. The two Romance languages show similar patterns, as they are modified around a third of their instances in all three verbs. In fact, the logistic regression analysis with the focus on modification, the model has shown to be significant in all four languages under study. The results for the variables analysed in this model are also in line with the findings from the thorough corpus-based analysis in previous subsections (cf. §3.4, §3.5, §3.6).

One of the strongly explicative variables is the LV itself: the LV *give* is less likely to be modified than *take* and *make* in English, Catalan and Spanish; while the German LV *geben* is more likely to be modified than the other two verbs.

Regarding the degree of frequency of LVCs in the corpora, the quantitative analysis of the results has shown contradictory effects on the results, especially in the *give*-LVCs. In this verb, the high frequent English *give*-LVCs and Spanish *dar*-LVCs are more often modified than medium and lower frequency LVCs, while the effect is less clear in the case of German *geben*-LVCs and Catalan *donar*-LVCs. The former shows more modification both in high-frequency and low-frequency than in the medium-frequency group, while the latter has predominant modification only in the low-frequency lemmas. In parallel, *take*-LVCs there is a difference between the two Germanic and the two Romance languages: English *take* and German *nehmen*-LVCs have a more predominant modification in high- and medium-frequency LVCs, while Catalan *prendre* and Spanish *tomar*-LVCs present more modification in the low-frequency group. Finally, *make*-LVCs show a tendency shared by the four languages under study: the low-frequency lemmas are the most often modified. In fact, the logistic regression model has confirmed that the low-frequency group is more likely to be modified in the German and Catalan dataset, while the English results are the opposite – they are less likely to be modified – and for the Spanish dataset the degree of frequency is not statistically significant.

With respect to the subclasses of modification, some divergences have emerged in the presentation of the quantitative results. In general, adjectival modification is the

main subclass of modification in all languages for the three verbs, with a significant exception of the Cat. *fer*-LVCs and Sp. *hacer*-LVCs ('make') where the adjectival modification is present in similar proportions than PP modification. For this verb, the results for Eng. *make* and Ger. *machen*-LVCs also show some lemmas that select PP modification more often than with the other two LVs analysed – although the general proportion is still much lower than adjectival modification, which is the most predominant for the two Germanic languages in all three verbs. The selection of PP modifiers in the nouns combined with *make*-LVCs has been found to be influenced by the selection of the NVE, and the transitive base verb of the eventive deverbal nouns (cf. §3.6).

Apart from the trends found crosslinguistically and within each language, when the focus is placed on specific LVCs, some patterns should be discussed. The corpus data has shown that the most frequent LVCs do not show signs of lexicalization (contra other previous corpus-based studies). It is only a handful of highly frequent LVCs that have been found to often co-occur with specific adjectives, as in Sp. *dar media vuelta* 'to turn around', in Eng. *take a deep breath*, and in Ger. *falsche Angaben machen* 'give false details'. However, a majority of particular LVCs, regardless of their degree of frequency, combine with a wide range of adjectives and the tendencies of co-occurring adjectives and nouns are very low.

Finally, some of these findings have brought up some relevant issues for theoretical linguistics in relation to the analysis of nouns in general and in LVCs in particular. For instance, the general tendency of English LVCs is to have modified NVEs, which was also found in previous studies. This has previously been attributed to the fact that the weight of the nominal is the main reason why LVCs coexist with the synthetic verb. However, the corpus findings have proven to diverge between the English language and the behaviour of LVCs in the other three languages: German, Catalan and Spanish. The coexistence of LVCs and their synthetic counterpart, more specifically their semantic correspondence, will be the focus of the case study presented in Chapter 4.

Moreover, the selection of determined NVEs and bare nouns has been related to the general tendencies of the languages involved, which is in line with the hypothesis that LVCs behave as traditional verbo-nominal constructions. Also, the modification of NVEs through quasi-argumental PPs or relative clauses has been detected more in *make*-LVCs, which select more deverbal nouns as their NVE with a transitive verbal base. The internal

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 3. Empirical study

structure of the noun, thus, has an influence on the overall structure of the whole construction, which is relevant for the syntax-semantics interface. These particularities of the nominal element in LVCs will be key in the theoretical chapters of this dissertation, more specifically in the typology proposal in Chapter 6.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

Chapter 4. LVCs and their synthetic counterpart

After the analysis of the empirical data in Chapter 3, the semantics of LVCs are discussed in this chapter, focussing on data from Catalan LVCs and their interchangeability with synthetic verbs. The reason for choosing Catalan LVCs is that no comparative study between analytic and synthetic verbs has been undertaken yet for this language.

First, §4.1 introduces the rationale for this chapter where the object of study is narrowed down, and previous studies are presented.

Second, the data retrieval, tagging and analysis are explained in §4.2.

Third, the case study constituting the basis of the chapter is presented in §4.3; the quantitative results of all verbs are first compared and then each verb is analysed more in-depth: the LV *donar* (§4.3.1), *prendre* (§4.3.2) and *fer* (§4.3.3).

Fourth, the statistical analysis of the quantitative results is presented through a logistic regression model (§4.4), where the most relevant variables and predictors of interchangeability are discussed.

Fifth, the presentation of the results is followed by the discussion of the patterns identified (§4.5) within a more qualitative approach: examples will be given of each category in different verbs, and a comparison with previous studies will be drawn. The discussion is twofold: first, all relevant variables are discussed (§4.5.1); and then the identified patterns are exemplified (§4.5.2).

The chapter ends with the conclusion and some final remarks in §4.6.

4.1. Introduction

As repeatedly pointed out in the literature on LVCs, most of these constructions have a corresponding synthetic verb that *a priori* expresses the same meaning. However, certain differences have been noticed, which challenge the question whether speakers convey the same meaning with the synthetic verb (e.g. *to answer*) and the LVC construction (e.g. *to give an answer*). Thus, the goal of this chapter is to first identify the differences between the LVC and its synthetic counterpart, as they have been reported in

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

the literature on these structures and then conduct a corpus-based comparative investigation of Catalan.

As presented in the introduction (Chapter 1), one of the most highlighted properties of LVCs is that they have a synthetic verbal counterpart with semantic proximity[32], as in (1).

(1) a. take a shower – to shower

   b. *Unterricht geben* (lit. ' lesson give') – *unterrichten* ('to teach')

   c. *prendre una decisió* (lit. 'take a decision') – *decidir* ('to decide')

   d. *dar ayuda* (lit. 'give help') – *ayudar* ('to help')

The paraphrasis is, however, not always a possibility and the equivalence between a LVC and a synthetic verb is not always present within a set language, although it may be found cross-linguistically (2).

(2) a. Eng. take advantage – Sp. *aprovechar(se)* ('to take advantage')

   b. Ger. *Abstand nehmen* (lit. 'take distance') – Eng. to reject

   c. Cat. *fer broma* (lit. 'do joke') – Sp. *bromear* ('to joke')

   d. Sp. *dar clase* (lit. 'give lesson') – Ger. *unterrichten* ('to teach')

For Piera & Valera (1999), there are four possibilities regarding the correspondence of the LVC and a synthetic counterpart, exemplified in (3): (i) the synthetic counterpart is equivalent in both its morphology and semantics (3a), (ii) there is morphological affinity but no semantic equivalence (3b), (iii) there is only a semantic equivalence (3c), and (iv) there is no synthetic counterpart within the language (3d).

---

[32] Actually, the lexicographic traditions in the different languages under study treat unergative verbs (e.g. the verb *to scream*) as equivalent to the LVC, as observed in Catalan (i), Spanish (ii) and German (iii), but it uses a different but synonymous noun in the *OED* for English (iv). Although this difference can be ascribed to the specific characteristics and tendencies in lexicographic traditions in the different languages, the correlation between the LVC and their synthetic counterpart is made clear with the inclusion of the paraphrases in the form of LVCs as the definition of synthetic verbs.
   (i) Cridar: 1 v. intr. [LC] Fer crits. (DIEC)
                 'do screams'
   (ii) Gritar: 2. intr. Dar un grito o varios.   (DRAE)
             'give a scream or more'
   (iii) Schreien: 1.a) einen Schrei, Schreie ausstoßen; sehr laut, oft unartikuliert rufen (DUDEN)
               'to produce a scream, screams; to cry very loud, often without articulation'
   (iv) Scream: Give a long, loud, piercing cry or cries expressing extreme emotion or pain. (OED)
In addition, in stylistic recommendations for administrative and legal documents (i.e. for Catalan, Bofarull & Montserrat, 2001; for Spanish, Montolío, 2011; Vilchez Vivancos & Sarmiento González, 2011), synthetic verbs are favoured over LVCs, the use of which is not recommended because nominalizations are penalized (Castillo, 2013).

(3) a. Ger. *Antwort geben* lit. 'answer give' - *antworten* 'to answer', Sp. *dar besos* lit. 'give kisses' – *besar* 'to kiss', Cat. *fer una còpia* lit. 'make a copy' - *copiar* 'copy'.

b. Eng. give voice – to voice 'to express', Sp. *hacer fiesta* lit. 'do holiday' – *festejar* 'celebrate', *donar propina* lit. 'give tip' - *propinar* 'to give/administer'

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

d. take advantage, *Abstand nehmen* 'distance take' – to reject, *donar una conferència* 'give a conference' – Sp. *conferenciar* , *dar clase* 'give class' – to teach

Although LVCs tend to correlate with a synthetic verb that expresses more or less the same meaning, as in (3a-c), the semantic equivalence between the pairs is not always complete. In some cases, the LVC expresses an action that is bound in time (4a), whereas its synthetic counterpart has a longer duration (4b) (Alonso-Ramos, 2004: 182).

(4) a. He gave a scream (*for two minutes).

b. He screamed for two minutes.

These aspectual differences are investigated by Bonial & Pollard (2020) for English LVCs on the basis of corpora. For some LVCs, the differences in telicity are clear, as in (5), where the synthetic verb does not have the endpoint (5a) that the LVC has (5b). However, this contrast is only found when the NVE is introduced by a determiner. Hence, the instances in (5c, d) show that if the nominal element is plural, telicity does not arise, which is further proved by its compatibility with the progressive LV, indicating that the event is durative rather than punctual (Van Valin & La Polla, 1997). Thus, there is no contrast between the synthetic verb *to think* and the LVC in plural *have thoughts*.

(5) 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]

d. I was just *having really, really bad thoughts* towards the enemy. [–telic]

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

Other LVCs in English are combined with nouns that are already inherently telic, as in (6) and (7), which do not show any aspectual contrast between the synthetic and analytic expressions.

(6) 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]

(7) a. She *appeared* with me on VH1 'Celebrity Rehab.' [+telic]

b. This afternoon, Bahrain's King Hamad *made a rare appearance* on television. [+telic]   (Bonial & Pollard, 2020: 15)

Even when the nominal element is a mass noun, like *advantage*[33], the LVC is not necessarily telic (8). For Bonial & Pollard (2020: 16), there is no consistent relation between the aspect of the event denoted by the noun, and the telicity of the whole construction.

(8) She *takes advantage* of the situation. [–telic] (Bonial & Pollard, 2020: 16)

Therefore, the aspectual differences in telicity between LVCs and the synthetic verb cannot be due to the nature of the whole construction, but they are rather linked to the syntactic and semantic properties of the nominal element: whether it is introduced by a determiner or not, whether it is singular or plural and whether it is count or mass, i.e. its mereological properties.

Another difference is found in the argument structure of both constructions. While in some LVCs the stimulus can be dropped (9a), its synthetic counterpart cannot omit it (9b) (Acedo-Matellán & Pineda, 2019: 194).

(9) a. Rosa tiene miedo (a la oscuridad)

Rosa has fear (of the darkness)

b. Rosa teme *(la oscuridad)

Rosa   fears   the darkness

'Rosa is afraid of the dark'

---

[33] *Advantage* belongs to group of nominals which can be count or mass depending on the presence of the determiner (as introduced in 3.2.3).

Another characteristic of LVCs that distinguishes them from their synthetic counterparts is that LVCs allow for a wider range of modification patterns than the full verb. They can take several adjectives (10a), and even a full relative clause (10b) as modifiers (Levin & Ström Herold, 2015: 8).

(10)   a. Take two deep breaths and relax.

b. She looked at him and gave him a smile that, though small, had gratitude in it.

In fact, the choice between the LVC and the corresponding synthetic verb "is not as arbitrary as it has been assumed in the literature" (Storrer, 2007: 186), but a thorough description of which factors favour one or the other type of construction is still lacking.

A first approach to the correlation between LVCs and their corresponding synthetic verb is presented by Storrer (2007) with a corpus-based study on German LVCs (or support verb constructions, as labelled in her study). When focusing on the synthetic verbs, less than 50% of the instances occurred in contexts compatible with the LVC (Table 53). Hence, LVCs seem to be semantically more specific, since they can only paraphrase the synthetic verb (or base verb, in Storrer's terms) in contexts where they share the specificity of the meaning conveyed.

Base	Instances	LVC paraphrase possible
absagen	210	63 (30%)
unterrichten	247	113 (46%)
helfen	214	77 (36%)
wirken	247	108 (43%)

Table 53. LVCs paraphrases of base/synthetic verbs (Storrer, 2007: 180)

Due to the semantic specificity of LVCs, they can be used to establish rhetorical relations with the synthetic verb, as in (11), where the synthetic verb (*unterrichten*) is replaced by the unambiguous and syntactically modified LVC in the second part of the sentence (*einen planvollen, geordneten, wissenschaftlichen Unterricht*).

(11)   Systematisch *unterrichten* heißt im Gerätturnen einen planvollen, geordneten, wissenschaftlichen *Unterricht* zu erteilen. (Storrer, 2007: 181) 'In apparatus gymnastics systematically teaching means giving a well-planned, well-ordered scientific instruction.'

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

The compatibility of LVC and the synthetic verb is not arbitrary, since it not only depends on the specific meaning of the LVC in comparison with the synthetic verb, but it also, and very especially, revolves around the discourse construction.

When the replaceability of LVCs by a synthetic verb is analysed, Storrer (2007: 182) finds that many of them could not be replaced with an appropriate synthetic verb even if their meaning corresponds to at least a specific context of the synthetic verb. She provides three different patterns for this tendency: (a) lexicalization, (b) discourse cohesion, and (c) modification of the noun.

Regarding lexicalization, LVCs tend to develop a specific subsense of the base verb, as in (12), where *Hilfe leisten* ('give help') refers to a special type of assistance provided in the context of gymnastics: 'holding or supporting the gymnast'[34].

(12)   Wie aber wird an sprunghohen Geräten, bei schwierigen Übungen und

Abgängen *Hilfe geleistet*?   (Storrer, 2007: 182)

'But how can assistance be provided for difficult exercises and landings from high parallel bars?'

These observations are also made for other languages, such as Italian (Bratánková, 2013), where some LVCs have developed a lexicalized meaning which has become independent from the synthetic verb: *fare festa* 'make party' vs. *festeggiare* 'to celebrate' or *dare peso* 'give weight' vs. *pesare* 'to weigh'. This is not the case of all LVCs, since some LVCs are still commonly used as an analytic variant of the synthetic verb (*fare una telefonata* 'make a call' vs. *telefonare* 'to call'), and they are considered to be a stylistic variant: the use of the LVC is prevalent where there is a qualitative denotation of the action through the modification of the noun.

In the cases where the NVE serves as an antecedent, it cannot be replaced by the synthetic verb, as in (13), without completely changing the syntax of the full sentence.

(13)   Den Kindern soll nur *Hilfe* geleistet werden, wenn sie *sie* benötigen. *Sie*

erfolgt durch Fragen (…)   (Storrer, 2007: 183)

---

[34] In Storrer's words: "Although the use of the base verb is possible and would be interpreted properly, the appropriate expression of this subsense is *hilfe leisten* and not *helfen*." (p. 182)

'One should only provide assistance to the children when they need it. It (this assistance) can take the form of questions (…)'

With regard to the modification of the NVE, Storrer (2007) finds that a considerable number of nouns are modified by adjectives in German LVCs, although not a majority of them (Table 54). When this happens, the option to find an equivalent adverb to modify the synthetic verb is not always a possibility and paraphrasis is not straightforward.

LVC	Instances	Modified by an adjective
Hilfe leisten	310	85 (27.4%)
Unterricht erteilen	122	34 (27.9%)
Wirkung ausüben	275	196 (71.3%)
Absage erteilen	82	42 (51.2%)

Table 54. NVE modified by adjectives (Storrer, 2007: 184)

Although Storrer's corpus is small, the LVC-synthetic verb pairings show that "their mutual substitution has in many cases negative effects on the cohesion of the discourse and/or requires considerable changes in the construction of the sentence. In most cases, the two types of constructions are not arbitrary stylistic alternatives that convey the same meaning" (Storrer, 2007: 185). In sum, they are interchangeable because their meaning overlaps, although the information structure of the discourse can be deeply affected by such a change.

In a similar line, Ronan (2019) studies the semantic and syntactic factors present in the use of LVCs and synthetic verbs in the variety of Late Modern Irish English and establishes that modification is the strongest variable to predict the choice of an LVC, although other factors are involved, namely semantic and aspectual properties of the construction. The variables which emerged for this variety of English are shared with the previous studies mentioned: a more specific meaning of some LVCs compared to the synthetic verb; some aspectual differences between the LVC and the verb; emphasis of the LVC through the focus fronting of the noun or the lower presence of passive voice with LVCs. However, the statistical analysis proves that the only significant variable for Late Modern Irish English is modification – which does not necessarily mean that it still holds for the present-day stage of the language.

The coincidence of the different semantic and syntactic variables influencing the correspondence of LVCs and their synthetic verbal counterpart is relevant for the analysis presented in the upcoming section (§4.2).

In previous studies on Spanish LVCs, De Miguel (2006: 1310) defends that LVCs in this language are not equivalent to the corresponding synthetic verb, and she argues that the semantic load of the synthetic verb is always heavier than the LVC. Sanromán Vilas (2009) argues that even when there is semantic equivalence between the pairs, the alleged equivalence only takes place at an abstract level.

For Spanish instances extracted from corpora, Sanromán Vilas (2009: 311) proves that LVCs and their synthetic counterparts are not interchangeable in many contexts. For instance, when the synthetic verb needs to specify the object, as the transitive verb *ordenar* 'to order' does in (14), the corresponding LVC allows the possibility of not mentioning the object at all, as in *dar una orden* 'give an order' because the NVE is the object of the construction.

> (14)   a. cuando un tío da una orden, se cumple a rajatabla.
>
>   b. cuando un tío ordena *(algo), se cumple a rajatabla.
>
> (Sanromán Vilas, 2009: 304)
>
> 'When a guy commands, instructions are followed to the letter.'

In other cases, LVCs appear to be more restricted than their verbal counterparts in some cases, such as with non-human subjects, as in (15), or in performative utterances, as in (16).

> (15)   a. *el viento daba golpes en ellas.
>
>   b. el viento golpeaba en ellas.
>
>   'The wind hit them.'
>
> (16)   a. ?Les doy la orden de que guarden silencio.
>
>   b'. Les ordeno que guarden silencio.   (Sanromán Vilas, 2009: 303)
>
>   'I command you to keep silent.'

However, the restrictions found in (15) and (16) should be taken cautiously, as examples in (17) and (18) indicate the opposite. There are corpus instances of the LVC *dar un golpe* 'to hit' with a non-human subject, as *el huracán Katrina* 'Hurricaine Katrina'

in (17a) and *la vida* 'life' (17b), and even with the same subject *el viento* 'the wind' in (17c).

(17)   a. En agosto de 2005, el huracán Katrina provocó un enorme desastre natural y fue el causante de la muerte de miles de personas, pero también *le dio un duro golpe* a la escena del jazz en Nueva Orleáns.

(CORPES XXI, 2006)

'In August 2005, Hurricane Katrina caused a massive natural disaster, killing thousands of people, but it also gave a heavy blow to the jazz scene in New Orleans.'

b. pero la vida *le dio un duro golpe* demasiado temprano.

(CORPES XXI, 2012)

'but life gave him a hard blow too early.'

c. El frío es intenso, el viento *da golpes helados* salpicando el rostro con polvo reseco.   (Web)

'The cold is intense, the wind gives icy blows splashing the face with dry dust.'

Also, some performative utterances can be found in online newspapers with LVCs, as *dar una orden* 'give an order' in (18a), and even in published books, as *hacer una promesa* 'make a promise' in (18b).

(18)   a. *te doy la orden* de que suspendas publicación hasta que yo conozca el libro y recibas mi autorización para publicarlo».   (La Razón)

'I give you the order to suspend publication until I know the book and you receive my authorization to publish it.'

b. si logro salir de ésta *te hago la promesa* de arrastrar un leño hasta tu altar   (Google Books)

'If I manage to get out of this, I promise to drag a log to your altar.'

Taking this into account, the variables will be included in our corpus-based study in order to establish whether they are relevant for the coexistence of LVCs and their synthetic counterparts.

With such a variety of contexts, it is necessary to formulate rules for the paraphrases between LVCs and their synthetic counterpart that specify the necessary conditions for their interchangeability. This is especially relevant for natural language

generation and processing, as well as for machine translation. As Bonial & Pollard (2020: 21) point out, there is a "simplifying assumption that all LVCs are interchangeable with a synthetic verb" which constitutes the basis of some natural language projects (such as the Abstract Meaning Representation project, Banarescu et al., 2013).

However, when investigating corpora, LVCs are constructions that do not fit an expected pattern. One of the main differences identified by Bonial & Pollard (2020) between LVCs and their synthetic counterpart is related to modification:

> Although there is some overlap in what can be expressed by nominal modification and what can be expressed by verbal modification (e.g., the manner of the event, deliberate vs. deliberately), much of the modification within LVCs has no counterpart in verbal modification, and could only be expressed periphrastically (p. 20).

Some cases of adjectival qualitative modification in LVCs cannot keep their original meaning when transformed into an adverbial modifier of the synthetic counterpart, because the modifier is ambiguous or does not sound natural, as in (19), whereas in other cases the adjective does not have an adverbial equivalent (20).

(19)　　a. We had – we had a *really good* laugh and then things happened.

　　　　b. [?]We laughed *really well* and then things happened.

(20)　　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.

<div align="right">(Bonial & Pollard, 2020: 20)</div>

In fact, LVCs allow the speakers to express the event in a noun and, thus, describe this event through the modification of the noun which is seen to be more flexible than verbal modification in different languages, such as German (Storrer, 2007) and English (Bonial & Pollard, 2020). The ease of modification of the noun in contrast to the adverbial modification of the synthetic verb has been considered one of the reasons for the high productivity of LVCs cross-linguistically (Huddleston & Pullum, 2002; Rácz et al., 2014).

With these grounds, Levin & Ström Herold (2015) examine the modification of LVCs in Germanic languages (English, German and Swedish), and they find that the most frequent modification of nominals in LVCs is adjectival modification. However, this conclusion needs a qualification. As described in Chapter 3, this is true for the two

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

Germanic languages in our corpora, but not for all LVs in the two Romance languages, where the PP modification is also relatively present.

The aim of the present chapter is to build on previous studies to advance in the description and analysis of the correlation between LVCs and their synthetic counterparts across languages. The correspondence and their lack of interchangeability have been studied in English (Bonial & Pollard, 2020), German (Storrer, 2007) and Spanish (Sanromán Vilas, 2009) on the basis of corpus data. However, there are no previous studies on Catalan LVCs and their equivalence with their synthetic counterparts. The main contribution of this chapter is thus to fill this empirical gap and present corpus-based results that can dialogue with and be contrasted to findings from other languages, as well as detect new lines for further research on Catalan LVCs.

4.2. Data retrieval and analysis

The data used in this chapter correspond to the LVCs extracted for the empirical basis of the present dissertation. That is, the LVCs extracted from the *Corpus Textual Informatitzat del Català* (CTILC) described and commented on in Chapter 3 for Catalan will serve as the basis for the study of the interchangeability of these constructions and their synthetic counterpart, in the specific corpus contexts. Therefore, from the basic repertoire of light verbs that is shared cross-linguistically (Butt 2010), the present chapter will deal with three Catalan LVs: *donar* 'give', *prendre* 'take' and *make* 'fer'.

The study is based on a random sample of LVCs obtained by searching for the collocates of the LV and the NVE. After manually excluding instances of non-light structures, for every LVC a sample of up to 200 occurrences were analysed, a total of 2,985 tokens. All examples were classified according to the following grammatical factors: (i) determination (yes/ no), (ii) modification (subclasses) and (iii) number of the NVE (singular/ plural).

After that, an extra parameter was added to the factors: equivalence between the LVC and its synthetic counterpart in the specific corpus context. The annotation was done manually and checked against native speakers' judgements[35]: instances were annotated for interchangeability/replaceability (yes/no), and an observations column served as

---

[35] The annotation was done in parallel by two native speakers of the Catalan language (the author of this dissertation and a MA student specialised in Catalan syntax). It was first done individually and afterwards confronted to decide on the final annotation.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

means to clarify the classification and identification of the patterns detected. Some observations were also taken (and checked) from the above-mentioned corpus-based studies (Storrer, 2007; Sanromán Vilas, 2009; Ronan, 2019; Bonial & Pollard, 2020), while others were added for this specific study (#3 in Table 55).

Linguistic areas	Patterns
Discourse	1. Performative utterances
	2. Coordination
	3. Subordination
Morphosyntax	4. Type of subject
	5. Determination of the NVE
	6. Modification of the NVE
Semantics	7. No semantic correlation
	8. Fixed expressions

Table 55. Summary of patterns identified

The specific LVCs analysed were presented in Chapter 3. Table 56 repeats the LVCs analysed organised according to the LV, their frequency in the corpus, the NVE they select, and the synthetic verb used in each case to annotate the interchangeability.

Verb	Frequency	NVE	Synthetic verb
Donar	high	suport ('support'), cop ('blow'), resposta ('answer'), volta ('stroll'), explicació ('explanation')	recolzar[36] ('to lean') colpejar, picar ('to hit') respondre ('to answer') voltar, girar ('to turn') explicar ('to explain')
	medium	consell ('advice'), permís ('permission')	aconsellar ('to advise') permetre ('to allow, permit')
	low	empenta ('push'), definició ('definition'), conferència ('conference'), ajuda ('help'),	empentar, empènyer ('to push') definir ('to define') *[37] ajudar ('to help')

---

[36] This equivalence is not accepted by the Catalan norm (DIEC; GIEC, 2016), as the verb *recolzar* has the literal meaning of *to lean on*, but the more abstract meaning *to give support* is only accepted with the LVC. Most speakers, however, do have this synthetic verb as an equivalence of the LVC, as the examples in (i-iii).

　　(i) amb la missió d'estudiar el problema i proposar possibles actuacions, que haurien de ser discutides, 'aprovades i **recolzades** pel Consell, (CTILC, 2019)
　　'with the mission of studying the problem and proposing possible actions, which should be discussed, approved and supported by the Council'
　　(ii) Els partits independentistes **recolzen** l'aturada, també els comuns, (RTVE)
　　'The pro-independence parties support the stoppage, as do the *Comuns*'
　　(iii) "però els alcaldes els **recolzen**", (Direct speech in Diari ARA)
　　'but the mayors support them'

[37] According to DIEC2, the verb *conferenciar* has the meaning of seeing someone to consult and discuss a certain issue (original: *veure's amb algú per consultar o discutir un assumpte determinat*), which does not correspond to the meaning of the LVC *donar una conferència* ('to give a conference').

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

		bufetada ('slap'), pallissa ('beating'), bes/ada ('kiss')	bufetejar ('to slap') apallissar ('to beat') besar ('to kiss')
Prendre	high	*decisió* ('decision'), *consciència* ('consciousness'), *nota* ('note')	decidir ('to decide') conscienciar(-se) ('to make aware of') anotar ('to note down')
	medium	*possessió* ('possession'), *iniciativa* ('iniciative'), *partit* ('party'), *precaució* ('precaution'), *molèstia* ('bother')	posseir ('to possess') * posicionar-se ('to take a stance') prevenir ('to prevent') molestar-se ('to bother')
	low	*distància* ('distance'), *determinació* ('determination'), *bany* ('bath'), *resolució* ('resolution'), *impuls* ('boost'), *represàlia* ('reprisal')	distanciar(-se) ('to distance, to grow apart') determinar ('to determine') banyar(-se) ('to bathe') resoldre ('to resolve') impulsar ('to propel') represaliar ('retaliate')
Fer	high	*Referència* ('reference'), *esforç* ('effort'), *feina* ('work/job')	referenciar ('to refer') esforçar(-se) ('to make an effort') treballar, feinejar ('to work')
	medium	*Ús* ('use'), *pregunta* ('question'), *visita* ('visit'), *viatge* ('trip'), *petó* ('kiss')	usar ('to use') preguntar ('to ask') visitar ('to visit') viatjar ('to travel') besar, petonejar ('to kiss')
	low	*Anàlisi* ('analysis'), *dibuix* ('drawing'), *fotografia* ('picture'), *abraçada* ('hug'), *aclariment* ('clarification')	analitzar ('to analyse') dibuixar ('to draw') fotografiar ('to photograph') abraçar ('to hug') aclarir ('to clarify')
* No synthetic counterpart found.			

Table 56. LVCs selection and their synthetic counterparts used in this study

Having presented the methodology for data retrieval and their classification, the next section presents the quantitative results extracted from the Catalan LVCs data and their interchangeability (or lack thereof) with synthetic verbs. The main quantitative analysis of the results will focus on descriptive statistics which mainly describe the data through counts and percentages. When comparing groups, some inferential statistics will

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

be introduced allowing to make inferences about the generalization of observed patterns, such as the chi-square test (already used previously in Chapter 3). Calculations have been carried out via MS Excel (through the Analyze Data toolbox), and the chi-square tests have been calculated on the Lancaster Stats Toolbox online (Brezina, 2018)[38].

4.3. Case study: the interchangeability of Catalan LVCs

This section will revolve around the quantitative results of the corpus-based study on the interchangeability of Catalan LVCs with synthetic counterparts. First, the results of all three verbs will be first compared. Afterwards, each verb will be further analysed: the LV *donar* in §4.3.1, *prendre* in §4.3.2 and *fer* in §4.3.3.

When the results are considered as a whole, table 57 shows that interchangeability of the LVC with its synthetic counterpart is a possibility in a good deal of cases (57.89%), while the contexts in which the LVC is not replaceable is also quite high (42.11%).

Interchangeability	Counts (%)
Yes	1728 (57.89%)
No	1257 (42.11%)
Total	2985

Table 57. General results on Catalan LVCs interchangeability

These findings present a general tendency but there is not a clear direction pointing towards the possibility of LVCs to be replaceable by a synthetic verb. Therefore, this could be one of the reasons for their crosslinguistic presence: they are needed, and do not contradict the economy of language principle (Zipf, 1949; Martinet, 1955).

Figure 46 shows that the tendency is homogeneous among the three verbs. In the case of *donar* 'give', the interchangeability trend is of 55.85% (versus 44.15% of non-replaceable cases). For *prendre* 'take', the results are a bit higher: 60.82% of instances are interchangeable with the synthetic verb, and 39.18% of instances are not. Finally, the verb *fer* 'make' shows results that are in the middle of the two previous verbs: 58.14% of instances are interchangeable with a synthetic verb, while 41.86% are not. This distribution is not statistically significant when all three verbs are compared ($\chi2(2)= 3.43$, p=0.1800479, Cramer's V = 0.034, negligible effect).

---

[38] Available at: http://corpora.lancs.ac.uk/lancsbox
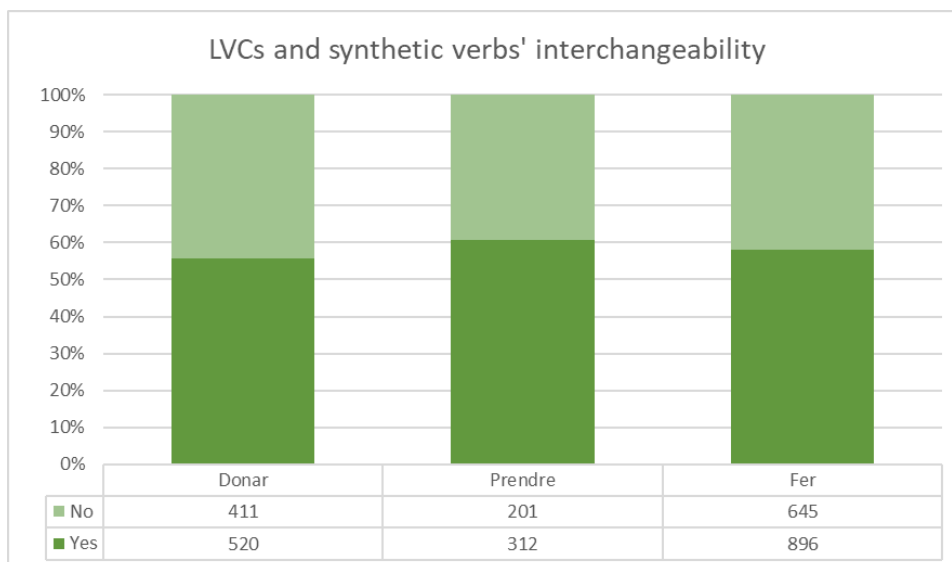
Figure 46. Distribution of Catalan LVCs interchangeability with synthetic verbs

The fact that the results for each verb in comparison does not present significant differences supports the general tendency found in table 57. More than half of the contexts where Catalan LVCs were found can be replaced with a synthetic verb, but this is only above half of the instances (55.18%). The contexts that cannot be replaced should, then, be further analysed in the following subsections.

4.4.1. Quantitative results: LV *donar*

A closer look at the correlation between the degree of frequency of LVCs in the corpora and synthetic interchangeability is needed to observe if frequency has an influence on the overall results. Figure 47 shows that the high frequent Catalan *donar*-LVCs can be interchangeable in the majority of instances (59.23%), which is similar to medium frequency (60.45%), while the low frequency LVCs are interchangeable in only 32.80% of cases. These results are statistically significant ($\chi2(2)=31.19$, p<.0001, Cramer's V=0.183, small effect).

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart
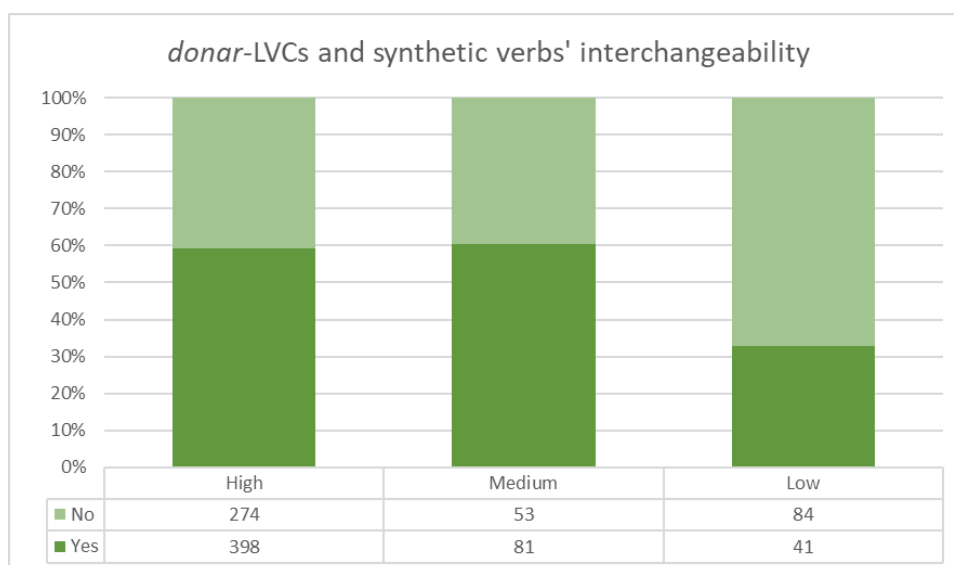
Figure 47. Distribution of Catalan donar-*LVCS interchangeability according to the degree of frequency (high, medium or low)*

Figure 48 presents the distribution of the interchangeability of Catalan *donar*-LVCs with a synthetic verb per lemma. An examination of each LVC shows that, indeed, the most frequent LVCs are mostly interchangeable in three cases: *donar suport* 'give support' (90%), *resposta* 'answer' (61.59%), and *volta* 'stroll' (70.37%), while the other two cases are less often replaceable: *donar cop* 'give blow' (35.65%) and *explicació* 'explanation' (16.96%). The results for medium frequency are more similar, and they are close to the general mean for the verb (55.85%, see Figure 46): *donar consell* 'give advice' (53.03%) and *permís* 'permission' (67.65%). Finally, the low frequency group is the most uneven because it includes LVCs with a small number of extracted instances. This groups involves LVCs which do not have a corresponding synthetic verb, as *donar conferència* 'give conference', and, thus, none of the instances are replaceable. It also contains two LVCs with low percentages of interchangeability, as *donar ajut/da* 'give help' (36.84%) and *definició* 'definition' (14.71%), but also other LVCs that are highly interchangeable: *bufetada* 'slap' (83.33%), *pallissa* 'beating' (66.67%) and *bes/ada* 'kiss' (70%).
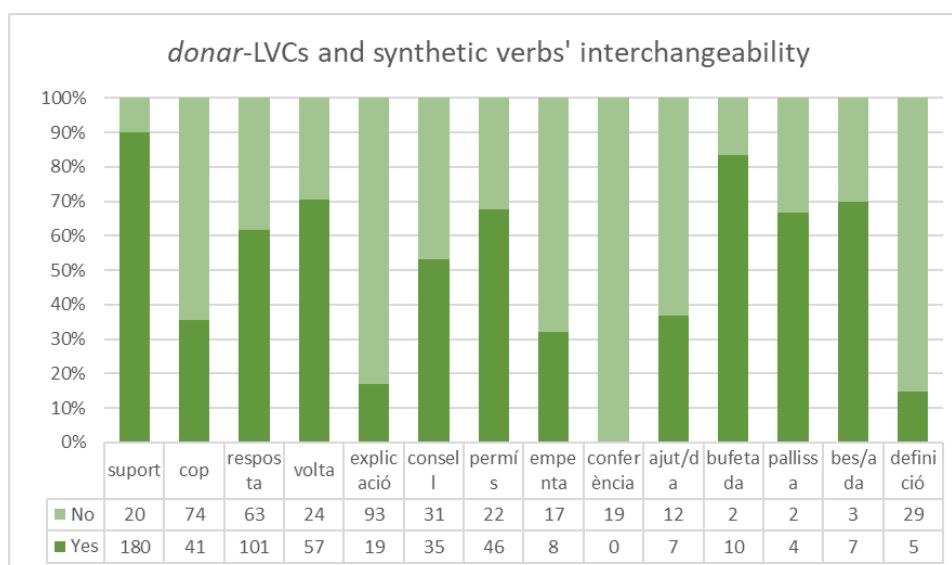
Figure 48. Distribution of Catalan donar-*LVCS interchangeability per lemma*

If the focus is placed on LVCs within the same semantic group, results also differ. For instance, the LVCs related to *beating* (that is, *donar cops* 'give blows'*, empenta* 'push' and *pallissa* 'beating') show very different results: while *donar cops* is replaceable in a third of the instances, the other two are above 65% of instances. Some of the patterns found in the instances that are not replaceable are related to the structure within the NVE: in (21a), the presence of the PP *de colze* 'of elbow' does not allow the equivalence with the synthetic verb *colpejar* nor *picar*[39]*;* in (21b)*,* the diminutive form *copets* 'little blows' involves a nuance that the verb does not hold; and in (21c), adjectival modification cannot be translated into an adverbial modification of the synthetic verb.

(21)   a. dins la capella els nens es podien *donar cops de colze*.

(CTILC, 2000)

'inside the chapel, the children could give each other elbow nudges'

b. i reien i es *donaven copets* al costat amb els colzes com si el joc les distragués de la feina.   (CTILC, 2003)

'and they laughed and gave each other little hits to the sides with the elbows, as if the game could distract them from the work'

c. tot *donant-hi grans cops* amb l'espadella per alliberar-ne les fibres interiors.   (CTILC, 2014)

---

[39] The verb *colpejar* 'to hit' is of infrequent use in present-day Catalan, while the LVC *donar cops* 'to give blows' is much more common. That's why the verb *picar* 'to hit' has also been considered as a possible verbal counterpart for this verb.

'while giving big blows with a small sword to release the inner fibbers'

In the case of *donar una empenta* 'give a push' and *pallissa* 'beating', as they are low-frequency LVCs, they do not represent many instances (12 and 6 respectively). Some of the cases cannot be replaced due to the presence of a diminutive form of the NVE, as in (22a), but the majority of examples show structural reasons for the lack of interchangeability. Specifically, the LVC is part of a relative sentence where the NVE belongs to a clause different from the LV, as in (22b,c).

(22)   a. va trobar una noia qualsevol, capaç de *donar-li l'empenteta* que encara li faltava.   (CTILC, 2002)

'he found a random girl, capable of giving him the little push that he still needed'

b. *l'empenta que va donar* Joan Antoni Samaranch a favor de la vela.

(CTILC, 2007)

'the push that Joan Antoni Samaranch gave in favour of sailing'

c. i els «veterans» em *van donar una pallissa* que em va deixar baldat.

(CTILC, 2011)

'and the veterans gave me a beating that beat me up'

In the case of *donar explicació* 'give explanation' or *donar definició* 'give definition', their interchangeability shows the lowest proportion. The reasons can be linked to previous observations in the literature, which show that it is due to sentence structure, such as the NVE being coordinated with another noun (23a), adjectival modification which cannot be translated into verbal modification (23b), and the fact that the noun and the LV belong to two different clauses in relative sentences (23c).

(23)   a. De seguida que *dono nom i definició* als meus sentiments,

(CTILC, 2004)

'As soon as I give name and definition to my feelings'

b. En català, el diccionari Fabra *dóna una definició senzilla*, però de caràcter més enciclopèdic que no pas lingüístic   (CTILC, 2005)

'In Catalan, the Fabra dictionary gives a simple definition, but with a more encyclopaedic approach than linguistic'

c. *La definició d'indignación que dóna* el Maria Moliner és la següent:

(CTILC, 2015)

'The definition of *indignación* that Maria Moliner's dictionary gives is the following'

d. *quines explicacions* els hauria pogut donar   (CTILC, 2008)

'what explanations I could have given them'

e. quan em vaig acomiadar, però aquesta vegada us hauria de *donar tantes explicacions* que me n'he sentit incapaç.   (CTILC, 2013)

'when I said goodbye, but this time I should give you so many explanations that I felt unable.'

If the modification and interchangeability properties of Catalan LVCs are correlated, Figure 49 presents a clear difference between modified and unmodified NVEs[40]. As discussed in Chapter 3, modification is not present in most instances of LVCs: Catalan *donar*-LVCs is only modified in 28.57%. However, when unmodified instances are compared with modified examples regarding their interchangeability, the divergences are wide: while only 29.70% of modified *donar*-LVCs are interchangeable, the proportion increases to 66.32% in unmodified cases. These results are also statistically significant ($\chi2(1)=103.31$, $p<.0001$, Cramer's V=0.333, medium effect).
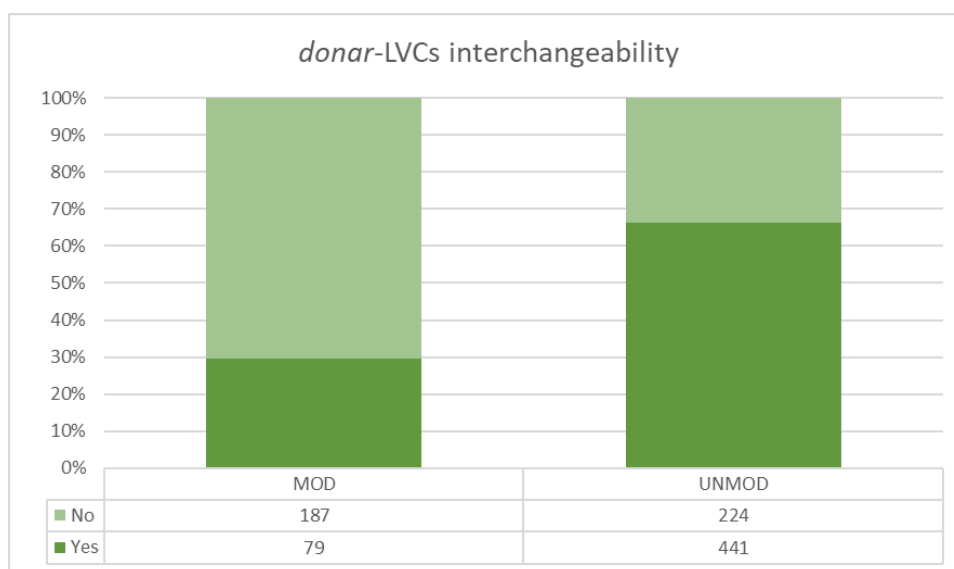
Figure 49. Distribution of Catalan donar-*LVCS interchangeability according to modification*

---

[40] All subclasses of modification described in Chapter 3 are included.

In sum, the analysis of *donar*-LVCs has shown certain tendencies which will need to be contrasted to the other two verbs. Regarding frequency, high and medium frequent LVCs are more interchangeable than low-frequency constructions. The fact that a LVC is more used in a language cannot be traced to a lack of corresponding synthetic. In fact, only one LVC does not have a synthetic counterpart, which is a low-frequency lemma: *donar conferència* 'give conference'. Also, the possibility to interchange is not homogeneous per lemma: some are highly interchangeable (e.g. *donar suport* 'give support', *bufetada* 'slap' and *bes/ada* 'kiss', are all interchangeable above 70% of instances), while others are not (e.g. *donar definició* 'give definition' and *explicació* 'explanation' show interchangeability below 30%). Finally, when the focus is placed on the reasons for the lack of interchangeability, the weight of the NVE and other structural properties can be pointed to. In fact, modification of the NVE –despite not being prevalent– does play a role in the probability of a LVC to be replaceable by a synthetic verb.

4.4.2. Quantitative results: LV *prendre*

In the case of *prendre*-LVCs, the correlation between the degree of frequency of LVCs in the corpora and their synthetic interchangeability shows little influence on the results. Figure 50 shows that high frequent Catalan *prendre*-LVCs can be replaced by a synthetic verb to a large extent (68.71%), followed by low frequency LVCs which are also replaceable in many cases (55.81%). In contrast, medium frequency LVCs are interchangeable just below half of the instances (48.99%). Although results are very similar, the differences are also statistically significant ($\chi2(2)=16.9$, p=0002, Cramer's V=0.182, small effect).
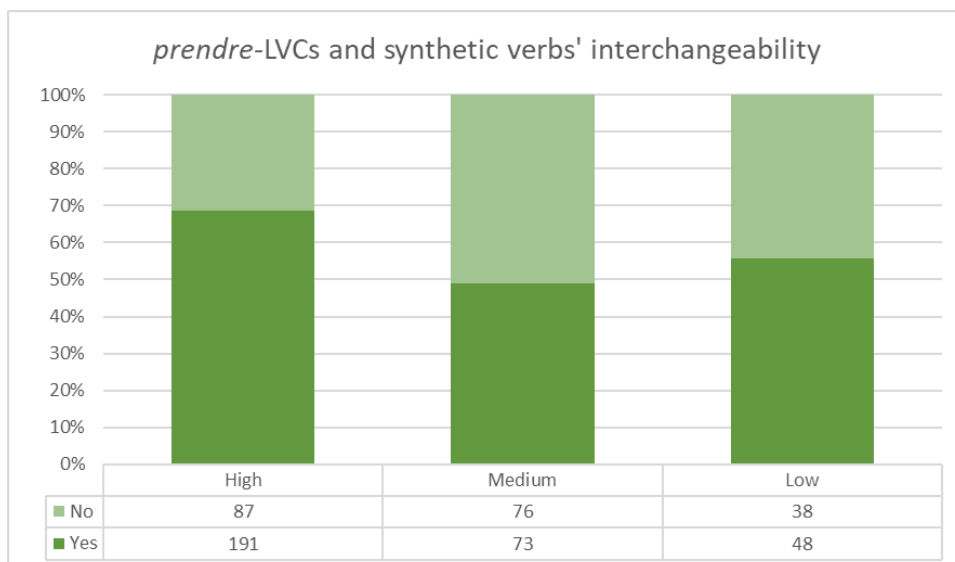
Figure 50. Distribution of Catalan prendre-*LVCS interchangeability according to the degree of frequency
(high, medium or low)*

Figure 51 presents the distribution of interchangeability of Catalan *prendre*-LVCs with a synthetic verb for each lemma. It is important to bear in mind that *prendre*-LVs are the ones that rank the highest in interchangeability (60.82% of instances are interchangeable with the synthetic verb, see Figure 46). In fact, the most frequent LVCs present high proportions of interchangeability by a synthetic verb: *prendre decisió* 'decision' (59.52%), *consciència* 'awareness' (89.52%) and *nota* 'note' (59.38%). This contrasts with the medium frequent LVCs, which is a more heterogeneous group: some LVCs are replaceable above the mean, as *prendre partit* 'take sides' (77.27%) and *molèstia* 'bother' (70.27%), while others are only interchangeable below half of the instances, as in *prendre precaució* 'take precaution' (46.15%) and *prendre possessió* 'take possession' (42.86%), and there is even one LVC which does not have a synthetic equivalent, *prendre iniciativa* 'take initiative'. Likewise, the low frequency group presents very different tendencies with four LVCs that are highly interchangeable: *prendre distància* 'take distance' (85.19%), *bany* 'bath' (66.67%), *impuls* 'boost' (66.67%) and *prendre represàlia* 'take reprisal' (75%); and only two with a very low proportion of interchangeability, as in *prendre determinació* 'take determination' (22.22%) and *resolució* 'resolution' (9.09%). As in the previous section, the low frequent LVCs display very short numbers of instances, because of their availability in the corpus and, thus, percentages should be taken cautiously.
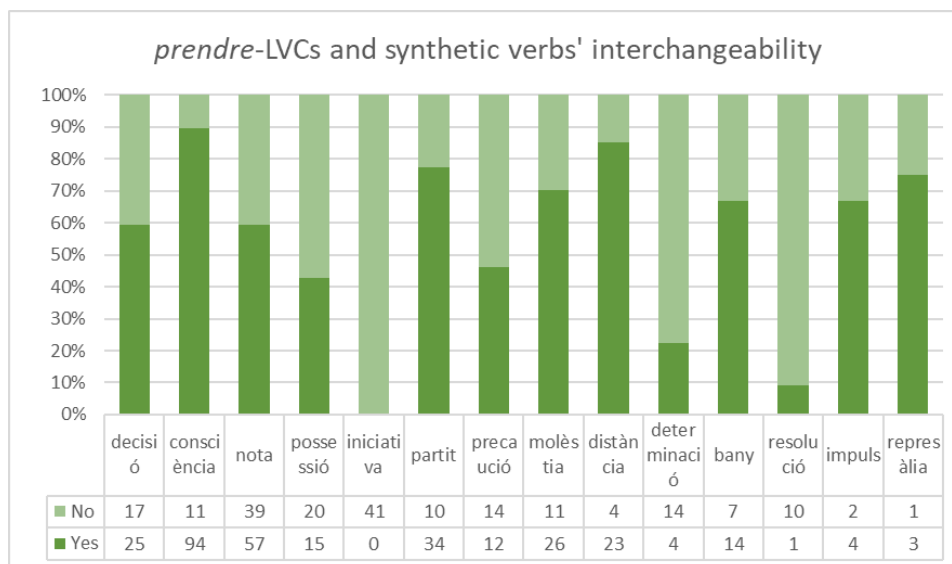
Figure 51. Distribution of Catalan prendre-*LVCS interchangeability per lemma*

Although this verb presents the highest interchangeability ratios, Figure 51 has shown that this tendency is highly lemma dependent. When the most interchangeable LVCs are considered, the most common patterns for non-replaceable instances are two: adjectival modification (24) and subordination (25).

(24)   a. cal prendre decisions coherents i sistemàtiques en fer la transcripció dels registres.   (CTILC, 2011)

'one must take coherent and systematic decisions when doing the transcriptions of registers'

b. Lo Pepe ara prenia consciència sobtada, violenta, que la gent de les idees del padrí era perseguida i detinguda.   (CTILC, 2000)

'Now Pepe took sudden, violent awareness that people with his grandad's ideas was followed and detained'

(25)   a. és una de les primeres decisions que cal prendre quan es vol crear una col·lecció.   (CTILC, 2009)

'it is one of the first decisions which one must take when creating a collection'

b. les notes que va prendre per a una conferència sobre la tauromàquia

(CTILC, 2008)

'the notes he took for a conference about bullfighting'

Likewise, the least interchangeable LVCs show similar structural patterns with the weight on the nominal element disallowing its interchangeability. Some elements in the

230

nominal domain are not straightforwardly replaceable in the verbal domain, such as the presence of a possessive determiner (26a), a demonstrative (26b) or adjectival modification (26c); and the subordination (27).

(26)   a. És cert que no han desaparegut del món, però Zeus ha pres *les seves precaucions*.   (CTILC, 2000)

'It is true that they didn't disappear from the work, but Zeus has taken his precautions'

b. A Lowood, vaig prendre *aquesta resolució*   (CTILC, 2001)

'In Lowood, I took this resolution'

c. Prenc *totes les precaucions possibles*.   (CTILC, 2000)

'I take all possible precautions'

(27)   amb *les coratjoses resolucions que havia pres* el consell de l'host.

(CTILC, 2005)

'with the courageous resolutions that the host council had taken'

Also, there is a case of lack of complete semantic correspondence with the synthetic verb in some instances of *prendre possessió* 'take possession', as in (28), where the synthetic verb *posseir* 'to possess' has the meaning of *possessing* which is not acceptable in the context of what is being possessed is a *position* in a work setting.

(28)   a. abans de *prendre possessió* del seu càrrec   (CTILC, 2005)

'before taking possession of his position'

b. *abans de posseir el seu càrrec

'before possessing his position'

If modification and interchangeability of Catalan LVCs are correlated, Figure 52 presents a difference between modified and unmodified NVEs. In this case, *prendre*-LVCs with modification in the NVE are interchangeable in just above half of the instances (51.30%), while unmodified corpus examples present higher results, over two thirds (64.90%). Although relatively close, these results are also statistically significant ($\chi2(1)=8.37$, p=0038, Cramer's V=0.128, small effect).
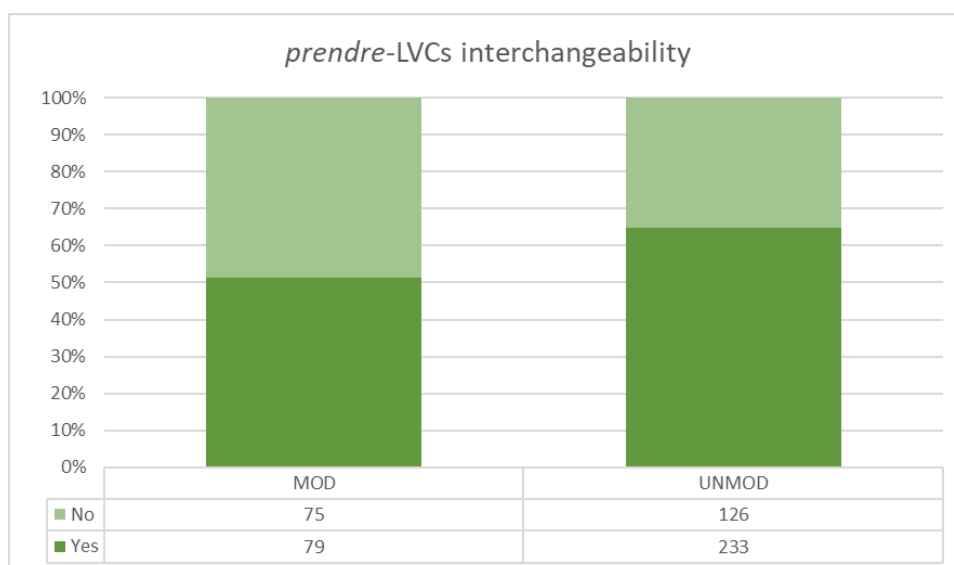
Figure 52. Distribution of Catalan prendre-*LVCS interchangeability according to modification*

In sum, frequency in *prendre*-LVCs show that high frequency LVCs are more interchangeable than the other groups. However, low-frequency LVCs are also highly replaceable, which should be taken cautiously because of its short number of instances. Among *prendre*-LVCs, there is only one LVC without a synthetic verb (that is, *prendre iniciativa* 'take initiative'), and there is another one which does not fully correspond semantically to its synthetic counterpart in every context (that is, *prendre possessió* 'take possession'). As already pointed out in Figure 51, the possibility to replace the LVCs with a verbal counterpart is not homogeneous when the lemmas are considered individually. Finally, in the patterns identified, the weight of the nominal element and the subordinated structures are the most relevant. Even so, the modified instances in *prendre*-LVCs are interchangeable in more than half of instances, but the proportion is lower than the cases when the LVCs are not modified.

4.4.3. Quantitative results: LV *fer*

In *fer*-LVCs, the correlation between the degree of frequency of LVCs in the corpora and its interchangeability shows similar results to the verb *donar*. Figure 53 shows that the high frequent Catalan *fer*-LVCs can be interchangeable in the majority of instances (61.37%), which is similar to medium frequency LVCs (59.24%), while low frequency LVCs are interchangeable in only 47.66% of cases. These results are also statistically significant ($\chi2(2)=14.47$, p=0007, Cramer's V=0.097, small effect).

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart
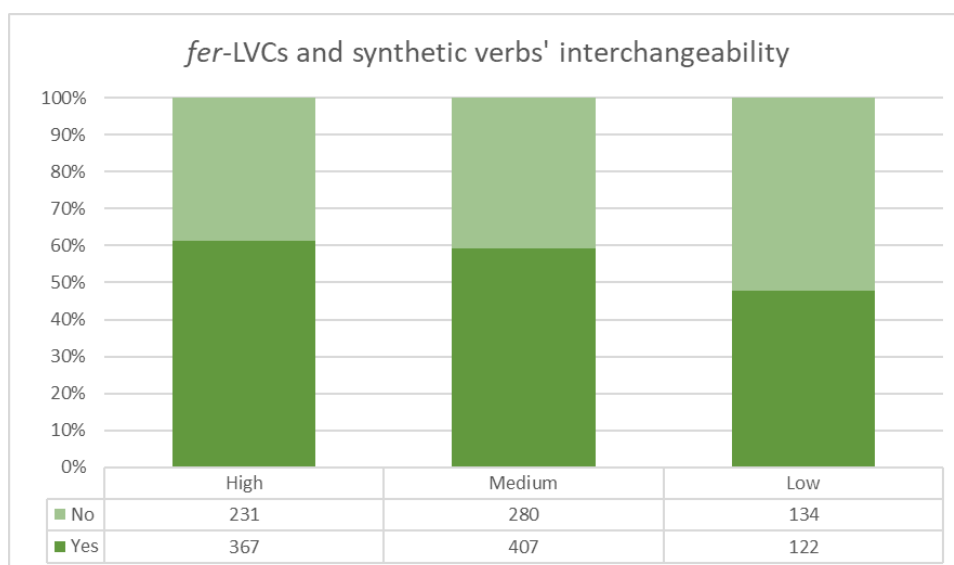
Figure 53. Distribution of Catalan fer-*LVCS interchangeability according to the degree of frequency (high, medium or low)*

Figure 54 shows the distribution of interchangeability of Catalan *fer*-LVCs with a synthetic verb for each lemma. The most frequent LVCs are highly interchangeable: *fer referència* 'make reference' (79.80%), *esforç* 'effort' (52.50%), *feina* 'work' (51.50%), but also some medium frequency LVCs: *fer ús* 'make use' (70.30%) and *petó* 'kiss' (83%). The other medium frequent LVCs are below half of the instances: *fer pregunta* 'make question' (41.22%), *visita* 'visit' (49.49%) and *viatge* 'trip' (42.95%). Finally, the low-frequency group presents the lowest proportion of interchangeability: *dibuix* 'drawing' (34.92%), *abraçada* 'hug' (37.50%) and *aclariment* 'clarification' (22.22%).
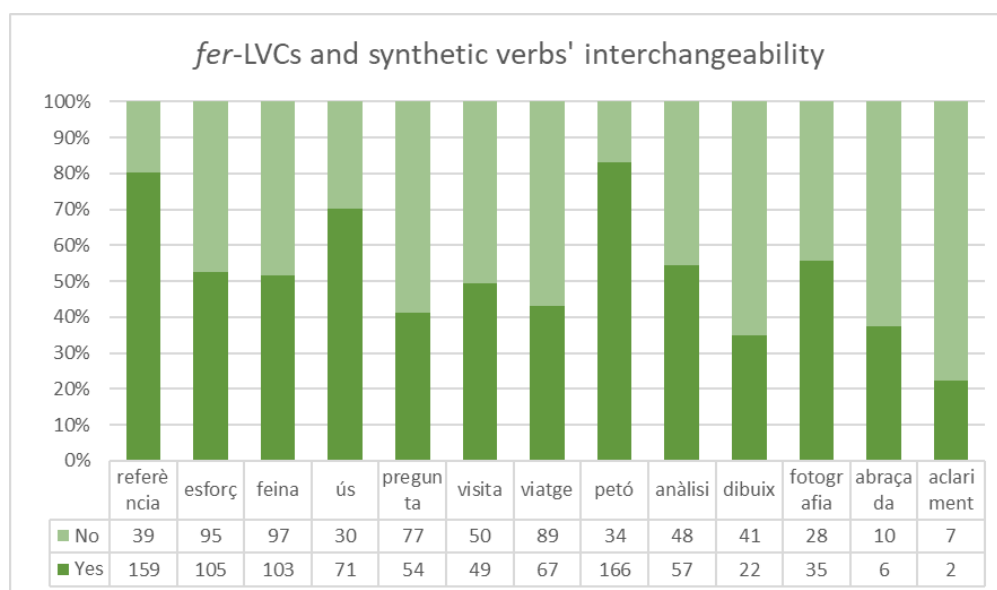
Figure 54. Distribution of Catalan fer-*LVCS interchangeability per lemma*

233

When the most interchangeable LVCs are observed, the most common patterns found in the contexts where the synthetic verb is not an option can be grouped in two: the weight of the nominal element (29) and sentence structure (30). For instance, in (29a,c) adjectival modification does not find an adverbial correlate, while (29b) the NVE is modified by a relative which disallows the substitution by a synthetic verb. Also, sentence structure influences the possibilities to be replaced by a synthetic verb, as in the case of subordination (30a) and the case of dislocation with the partitive *en* (30b).

(29)   a. A tal efecte, es *farà expressa referència* al model de gestió

(CTILC, 2001)

'To this effect, explicit reference will be made to the management model'

b. Sense *fer cap referència* que no fos sobreentesa als diners

(CTILC, 2004)

'Without making any reference to the money that was not tacit'

c. on es *fa un ús intel·ligent i saludable* del sentit de l'humor.

(CTILC, 2004)

'where an intelligent and healthy use of sense of humour is made'

(30)   a. Aquesta superioritat que us atribuïu dependrà de *l'ús que hàgiu fet* del vostre temps i de la vostra experiència.   (CTILC, 2001)
'This superiority that you attribute yourself will depend on the use that you have made of your time and experience'

b. I ara permeteu-me que *en faça ús*, de la potestat de vulnerar el mecanisme del temps   (CTILC, 2005)
'And now allow me to make use of it, of the authority of vulnerating the time mechanism'

In the least interchangeable LVCs (*fer aclariment* 'make clarification' and *dibuix* 'drawing')*,* the patterns revolve around the same aspects: the weight on the nominal element, such as adjectival modification in (31), and sentence structure, such as topicalization in (32a) or the morphological passive in (32b).

(31)   a. una cambra fosca permetia *fer dibuixos molt realistes*, (CTILC, 2011)
'a dark chamber allowed to make very realistic drawings'

b. Hem de fer aquí un breu aclariment sobre una inscripció en lletra minúscula i corresponent a l'any 1561. (CTILC, 2009)

'We have to make here a brief clarification about an inscription in lower case letters from 1561'

(32) a. Aquest dibuix el va fer la Gritli Moser —va explicar Matthäi—.

(CTILC, 2011)

'This drawing [it] was done by Gritli Moser – Matthäi explained —.'

b. es pot veure que el dibuix preliminar va ser fet a mà lliure en un color ocre vermell a la capa superficial del morter. (CTILC, 2012)

'it can be seen that the preliminary drawing was made by free hand in an ochre red colour on the superficial layer of the mortar'

When the modification and interchangeability of LVCs are correlated, Figures 55 shows a clear difference between modified and unmodified NVEs. For *fer*-LVCs, instances with modification in the NVE are interchangeable in only 39.72% of cases, while unmodified examples present a much higher proportion of results, well above two-thirds (69.08%). These results are also statistically significant ($\chi2(1)=17.57$, $p<.0001$, Cramer's V=0.288, small effect).
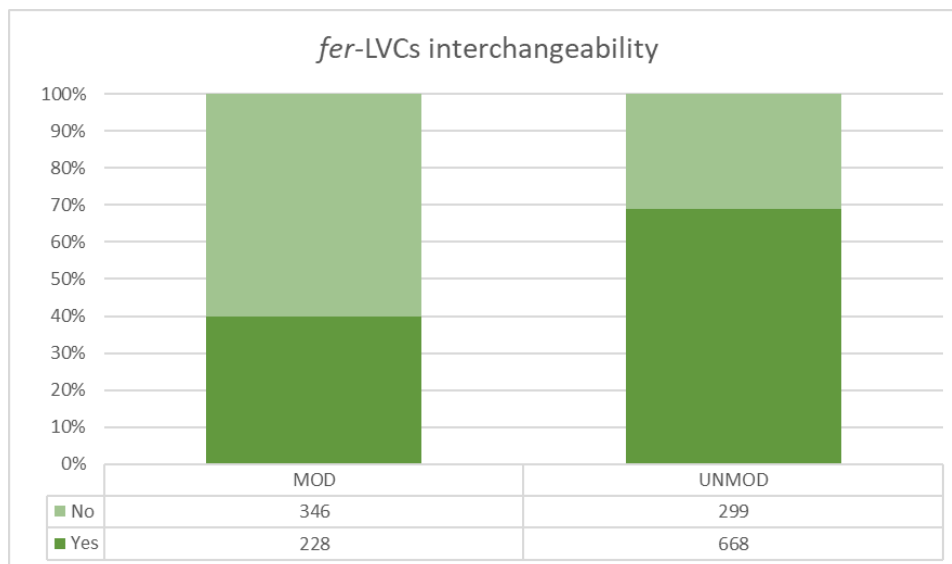
Figure 55. Distribution of Catalan fer-*LVCS interchangeability according to modification*

In sum, *fer*-LVCs show similar tendencies as the previous two LVs with some differences in the percentages. Basically, high-frequency *fer*-LVCs are the most interchangeable, while low-frequency LVCs are the least. Also, the patterns identified for the lack of interchangeability with certain lemmas can be explained in relation to two

factors: the weight of the NVE through modification, and sentence structure in contexts of subordination, passivization or topicalization. Hence, modification has proven to be a significant factor influencing the lack of replaceability of *fer*-LVCs.

4.4. Statistical analysis

In order to confirm the trends attested in the previous sections, I have employed the *binary regression analysis* to explore the data, a statistical method that uses explanatory variables to estimate their effect on the linguistic variable. As in the statistical analysis and discussion in Chapter 3 (§3.7), this analysis is implemented to observe the relation between the response variable (in this case, the interchangeability) and the other variables examined (that is, the predictors).

Binary logistic regression is chosen for the analysis of the object of study of this corpus-based research since the response variable (interchangeability) is not numerical and has only two possible values (that is, yes/no or interchangeable/not interchangeable). The purpose to implement this statistical test is to examine the extent to which the morphosyntactic variables considered (number, determination and modification) as well as the most general variables (frequency) can predict or explain the patterns of modification attested.

In logistic regression, we look at both the general performance of a model as well as at individual coefficients showing the effect of the predictor variables on the outcome of interest. 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"[41].

Table 58 shows that the logistic regression model was statistically significant, Likelihood ratio test (LL): 677.61 (p < 0.0001). The model explained 27% (Nagelkerke R2) of the variance in interchangeability. Since the outcome variable is interchangeability (yes/no), the first thing to bear in mind is that the reference level of the response variable is the first one: lack of interchangeability (a_no). This is important as the coefficients of the predictor variables (Det, Mod, Num, Freq) have to be interpreted in relation to the second level of the response variable, i.e. interchangeability (b_yes).

---

[41] Available at: http://corpora.lancs.ac.uk/lancsbox

Chapter 4. LVCs and their synthetic counterpart

Logistic Regression Model

Overall model statistics:
Overall significance: Likelihood ratio test (LL): 677.61 (p < 0.0001)
Model accurary: C-index: 0.76
Predictive power: Nagelkerke R^2: 0.27
AIC: 3397.85

Coefficients:

	Estimate (log odds)	Standard Error	Z value (Wald)	p-value	Estimate (odds)	95% CI lower	95% CI upper
(Intercept)	-0.711	0.091	-7.790	0.000	0.491	0.410	0.587
Detb_bare	1.665	0.099	16.746	0.000	5.286	4.359	6.437
Modb_unm	1.104	0.087	12.669	0.000	3.016	2.544	3.580
Numb_plur	-1.484	0.105	-14.184	0.000	0.227	0.184	0.278
Freqb_medium	0.224	0.095	2.366	0.018	1.251	1.040	1.506
Freqc_low	-0.049	0.120	-0.410	0.681	0.952	0.753	1.204

Table 58. Output of the logistic regression model

Let us now briefly comment on the significance of the variables and the interpretation of their coefficients. First, the p-value will be taken into account when commenting on the relevant variables (≤0.05), and it will be combined with the estimate (log odds) which expresses the odds or probability of the specific variable to be interchangeable. In this case, the variable Det=b_bare shows a positive estimate (log odds), which indicates that it significantly increases the odds of Inter=yes compared to the reference level of this variable, i.e. Det=a_Det (which are NVEs introduced by a determiner). More specifically, bare nouns are 5.286 more likely (estimate, odds) than determined LVCs to be interchangeable. Likewise, unmodified nominals (Mod=B_unm) show a positive estimate, and they are 3.016 more likely than modified nouns to be interchangeable. Finally, plural NVEs (Num=b_plur) show a negative estimate and, thus, are less likely than singulars to be interchangeable, which is in this case 0.227 lower.

In contrast, frequencies also show relevant results but to a lesser extent, as it is only statistically significant in the case of medium-frequency LVCs (p=0.018). In this

case, Freq=medium shows a positive estimate which indicates that it is more likely than the reference variable (Freq=high) to be interchangeable, more specifically 1.251 more likely.

This statistical analysis is relevant because all variables have been compared in relation to the response variable (that is, interchangeability). These statistical results will be discussed in the upcoming subsection and will be contrasted to previous analyses.

4.5. Discussion

The discussion of the results is organised according to the different variables that have been taken into account in this study (§4.5.1) with the aim to compare them with previous studies. After that, the patterns found will be commented with particular corpus instances (§4.5.2).

4.5.1. Relevant variables

The first relevant variable is the LV, which was not included in the logistic regression analysis because it was proven to be not statistically significant with the chi-square test (see Figure 46). In this case, the three verbs under study (*donar* 'give'*, prendre* 'take'*, fer* 'make') have presented similar results with respect to their interchangeability, and no relevant divergences were detected. Unfortunately, no previous studies have been found where a comparison between different verbs has been put forward in this regard.

A second variable is the frequency of LVCs within the corpus. A first approach to this issue has shown that, in the three verbs under study, low-frequency LVCs tend to be the least interchangeable. This variable, however, has not shown to be statistically significant in the logistic regression analysis (see Table 58 above), which could be due to the fact that the extracted instances for this group are very low. In the statistical analysis, the medium frequency group proves to be more interchangeable than the high frequency group. The relevance of these findings is that the medium frequent LVCs are more interchangeable with the synthetic than the other groups, which could seem counterintuitive with the idea that LVCs should express something different than the base verb in order to be productive in the language. Again, no previous studies have focused on the frequency of LVCs when analysing their interchangeability, and this is, thus, a main contribution of this study.

The remaining morphosyntactic variables have been previously approached in studies on LVCs, that is, determination, number and modification. In the case of

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

determination, the statistical analysis has shown that bare nouns in Catalan LVCs favour interchangeability which is in line with previous findings (Mendívil, 1999; Alonso-Ramos, 2004, for Spanish). Also, when number is taken into account, plural NVEs are less probable to be interchangeable than singular NVEs. In this case, however, the intersection between bareness and number could not be included and the category bare nouns involves both bare singular and bare plurals. Hence, no generalisations in this line can be made.

Likewise, an analysis of modified and unmodified NVEs has pointed to a preference for interchangeability in the case of nominals without modification (see Figures 49, 52 and 55). Also, the logistic regression analysis has shown that, in fact, unmodified nominals are more likely interchangeable than modified nouns. These findings are also in line with previous corpus-based research findings across languages (Storrer, 2007; Alonso-Ramos, 2004).

Although some variables clearly favour interchangeability of Catalan LVCs with a synthetic verb, a more qualitative analysis of the identified patterns can contribute to the explanation of the statistical tendencies found. I turn to this qualitative analysis in §4.5.2.

4.5.2. Identified patterns

When the identified patterns were first classified (see Table 55 above), they were organised in three thematic areas: discourse, morphosyntax and semantics, because previous corpus-based studies have found tendencies related to these three linguistics areas (as in Bonial & Pollard, 2020, for English LVCs; Storrer, 2007, for German LVCs; and Sanromán Vilas, 2009, for Spanish LVCs). The fact that not all instances from the present corpus sample could be coded for one of these patterns led us to a more qualitative approach to their analysis and discussion.

The first thematic area is related to discourse. More specifically, three different patterns have been identified in this domain: performative utterances, coordination between nominals and subordinating structures.

Regarding performative utterances, only commands have been detected in the corpus sample (33), and they are all interchangeable, as illustrated in (34).

(33)　　a. —*Faci un esforç* —insisteix el comissari—, perquè és obvi que algú el

vol incriminar. (CTILC, 2018)

'—Make an effort —the Inspector insists —, because it is obvious that someone wants to incriminate you'

b. No *em donis empentes*. (CTILC, 2009)

'Don't give me pushes'

c. Govern d'Aragó, polítics d'Aragó: *preneu-ne nota*. (CTILC, 2009)

'Government of Aragon, politicians of Aragon: take note of it'

(34)   a. —*Esforci-s'hi* — insisteix el comissari—, perquè és obvi que algú el vol incriminar.

'—Make an effort —the Inspector insists —, because it is obvious that someone wants to incriminate you'

b. No *m'empentis.*

'Don't push me'

c. Govern d'Aragó, polítics d'Aragó: *anoteu.*

'Government of Aragon, politicians of Aragon: write down'

Therefore, the results from Catalan LVCs contrast with previous findings in Sanromán Vilas (2009) for Spanish LVCs. Our corpus sample for Catalan LVCs has provided examples of performative utterances performed through an LVC, and they are all interchangeable with the synthetic verb. In both cases, they hold the performative force.

With regard to coordination, several instances have shown to be not replaceable because the NVE within the LVC was coordinated with another noun. In this case, interchangeability is only possible when the two nouns are replaced by different verbs, as shown in (35) and (36).

(35)   a. L'escriptora i periodista no deixava de *fer petons i abraçades*.

(CTILC, 2013)

'The writer and journalist did not stop giving kisses and hugs.'

b. L'escriptora i periodista no deixava de *petonejar i abraçar*.

'The writer and journalist did not stop kissing and hugging.'

(36)   a. De fet, prova de la frivolitat amb què ens venen les xifres és que a l'hora de *fer estudis i anàlisis* s'accepta la tesi oficial. (CTILC, 2015)

'In fact, proof of the frivolity with which we are sold the numbers is that, when making studies and analyses, the official thesis is accepted'

b. De fet, prova de la frivolitat amb què ens venen les xifres és que a l'hora d'*estudiar i analitzar* s'accepta la tesi oficial.

'In fact, proof of the frivolity with which we are sold the numbers is that, when studying and analysing, the official thesis is accepted'

Also, the cases where the NVE and the LV are in different clauses due to subordination are not interchangeable because the synthetic counterpart cannot be divided in two clauses. As exemplified in previous sections, these cases are found in the three LVs under study, as illustrated here under (37).

(37)   *Els cops que dono* a la fusta amb la pistola em retrunyen fins a la ferida,

(CTILC, 2018)

'The blows that I give to the wood with the weapon resound until the wound'

Likewise, there are instances in which the LVC cannot be replaced by a synthetic verb because the NVE is the antecedent of another element in the discourse, as the accusative pronouns *els, les* and the partitive *en* (38a, b, c), as well as the demonstrative determiner *aquest* 'this' (38d). The replacement by a synthetic verb loses this reference across the discourse.

(38)   a. i si cal que fem esforços, *els* farem».   (CTILC, 2015)

'and if it is needed that we make efforts, we will do them'

b. Passats els sis mesos hi tornà, es féu noves anàlisis i *les* portà al doctor.

(CTILC, 2010)

'After six months s/he went back, made new analyses and brought them to the doctor's'

c. tant podien fer un viatge com fer-ne *dos o tres*. (CTILC, 2003)

'they could make one trip or two or three'

d. Ara em cal fer el viatge de debò, que és *aquest* de construir un món en aquesta vall, sentint com passa el dia a dia.   (CTILC, 2001)

'Now I must make the real trip which is that of building a work in this valley seeing how everyday passes'

The second thematic area is related to the morphosyntax of the LVC. This area includes issues on the type of subject, determination and modification of the NVE. First, lack of subjects in impersonal constructions appear as a hindrance to the interchangeability of the LVC with their synthetic counterpart, as in (39).

241

(39)   a. No va ser fins ben entrat el segle XX que *es va prendre consciència* que aquesta mancança era un escàndol   (CTILC, 2007)

'It was not until the 20th Century that awareness was risen of the fact that this shortage is a scandal'

b. Això és molt normal en ciència quan *es fa referència* a experiments o observacions en les línies més frontereres de la recerca. (CTILC, 2006)

'This is very usual in Science when a reference is made to experiments or observations in the closest lines to research'

In fact, if these impersonal constructions include a specific subject, the replacement with a synthetic verb is possible, as illustrated in (40).

(40)   a. No va ser fins ben entrat el segle XX que *ens vam conscienciar* que aquesta mancança era un escàndol

'It was not until the 20th Century that we became aware of the fact that this shortage is a scandal'

a'. No va ser fins ben entrat el segle XX que *hom es va conscienciar* que aquesta mancança era un escàndol

'It was not until the 20th Century that one became aware of the fact that this shortage is a scandal'

b. Això és molt normal en ciència quan *ens referim* a experiments o observacions en les línies més frontereres de la recerca.

'This is very usual in Science when we refer to experiments or observations in the closest lines to research'

b'. Això és molt normal en ciència quan *hom es refereix* a experiments o observacions en les línies més frontereres de la recerca.

'This is very usual in Science when one refers to experiments or observations in the closest lines to research'

Some issues regarding the characteristics of the subject were first brought up for Spanish LVCs by Sanromán Vilas (2009). She argues that non-human subjects are not compatible with LVCs, especially those with the Spanish LV *dar* 'give', because the subject is the agent of the event expressed by the LVC. However, some examples extracted from CORPES XXI and the Internet have proven otherwise, as seen above in (17). Likewise, in the present case study, no limitations related to the properties of the

subject have been found to influence the interchangeability of Catalan LVCs. Even when the action is involuntary, the instances are replaceable by the synthetic verb, as in (41).

(41) a. Què són aquests escarafalls! Que *t'has donat un cop a l'ull*? No t'entenc.

(CTILC, 2000)

b. Què són aquests escarafalls! Que *t'has colpejat l'ull*? No t'entenc.

'What is all this fuss! Did you hit your eye? I don't understand you'

Regarding the weight of the nominal element, both modification and determination have been found to have an effect on the interchangeability of Catalan LVCs. First, the presence of some determiners can hinder the interchangeability possibilities when it is specific, like numerals (42a, b) or demonstratives (42c).

(42) a. Es prenen igualment *tres* decisions importants:

(CTILC, 2010)

'Three importants decisions are made anyways'

b. A la jornada d'ahir, Gené només va poder donar *24* voltes

(CTILC, 2001)

'In yesterday's session, Gené could only make 24 turns'

c. Lorena, dóna *aquesta* resposta   (CTILC, 2015)

'Lorena, give this answer'

Also, modification has shown to play a role in the weight of the nominal element and the possibilities of LVCs interchangeability, as it has also been proven in the statistical analysis in §3.7. The relevance of adjectival modification in contributing to the semantics of the event expressed by the LVC has been widely studied: it has generally been proposed that the adjectival modification in LVCs can be expressed through adverbial modification in the use of synthetic verb (i.e. Alonso-Ramos, 2004: 200), unless there is no equivalent adverb (Cattell, 1984: 8). This is the case of most instances reported in previous sections, as repeated here in (43).

(43) a. Prenc *totes les precaucions possibles*.   (CTILC, 2000)

'I take all possible precautions'

a'. *Previnc tot possiblement.

'I prevent all possibly'

b. una cambra fosca permetia *fer dibuixos molt realistes*,

(CTILC, 2011)

243

'a dark chamber allowed to make very realistic drawings'

b'. *una cambra permetia dibuixar realistament

'a dark chamber allowed to draw realistically'

However, adjectival modification is not the only class of modification included in the tagging. Other kinds of modifications can also hinder the possibilities of interchangeability, such as comparative structures (44a), relative complementation (44b), and even the diminutive form of the NVE (44c).

(44)   a. L'oracle se n'estarà prou de *donar una resposta tan clara com la pregunta*,   (CTILC, 2000)

'The oracle will avoid giving an answer as clear as the question'

b. i que al mateix temps *es donessin respostes que regulessin la conflictivitat social en el territori.*   (CTILC, 2010)

'and that at the same time answers were given that would regulate the social potential for conflict in the territory'

c. Doctor Hochroitzpointner, passi, passi a *donar una volteta* per aquí.   (CTILC, 2015)

'Doctor Hochroitzpointner, come in, come in, to take a little stroll around here'

All these instances would lose the meaning introduced by the modification when the LVC is replaced by a synthetic counterpart, as illustrated in (45).

(45)   a. L'oracle se n'estarà prou de [?]*respondre tan calarament com la pregunta*,

'The oracle will avoid answering as clear as the question'

b. i que al mateix temps *es respongués *que regulessin la conflictivitat social en el territori*.

'and that at the same time it was answered that would regulate the social potential for conflict in the territory'

c. Doctor Hochroitzpointner, passi, passi a [?]*voltar poquet* per aquí.

'Doctor Hochroitzpointner, come in, come in, to stroll a little around here'

Finally, the third thematic area focuses on the semantics. As already established in Table 56 above, some LVCs lack a synthetic verb that could be used to replace the construction (i.e. *donar conferència* 'give conference' and *prendre iniciativa* 'take initiative'). In addition, there are some LVCs with more than one verb for its replacement,

244

e.g. *donar volta* which can be replaced by *voltar* 'go around' and *tomar*, as in (46), and *girar* 'turn around' (47).

(46) a. *vaig donar una volta* pel claustre, em vaig detenir a llegir el que hi havia escrit en alguns taulons d'anuncis,   (CTILC, 2010)

'I took a stroll around the cloister, I stopped to read what was written in some noticeboards'

b. *vaig voltar / tombar* pel claustre, em vaig detenir a llegir el que hi havia escrit en alguns taulons d'anuncis,

'I walked around the cloister, I stopped to read what was written in some noticeboards'

(47) a. Quan la mare *va donar la volta* per marxar, va dir dues o tres vegades:

(CTILC, 2003)

'When the mother took a turn to leave, she said twice or three times:'

b. Quan la mare *es va girar* per marxar, va dir dues o tres vegades:

'When the mother turned to leave, she said twice or three times:'

Even when the semantic correlation between the LVC and the synthetic verb is equivalent, other semantic parameters can influence their interchangeability. One is the abstract meaning of LVCs in certain contexts, which do not hold when the synthetic verb is used, as in *donar l'empenta definitiva* 'give the definitive push' to a project in (48a) which cannot be replaced by *empentar* 'to push' or *prendre nota* 'take note' about someone's presence in (48b) which is not interchangeable by *anotar* 'to write down' because the event is not literal.

(48) a. Fou Enric Prat de la Riba, en accedir a la presidència de la Diputació de Barcelona, qui *va donar l'empenta definitiva* al projecte: (CTILC, 2003)

'It was Enric Prat de la Riba, in accessing the Presidency of the Diputació in Barcelona, who gave the final push to the project'

b. Zoltán Vègh *va prendre nota* de la presència de l'argentí Miguel Escamilla Stangel al despatx del governador provincial. (CTILC, 2012)

'Zoltán Vègh took note of the presence of the Argentinian Miguel Escamilla Stangel in the provincial governor's office'

Finally, there are fixed expressions with some LVCs which have lexicalized a specific meaning that cannot be interchanged with the synthetic verb. In the present

sample, however, these cases are a minority and they have been detected for very specific contexts, as in *donar la volta al món* 'go around the world' (49a) or *prendre nota* 'take the order' in restauration contexts (49b).

(49)   a. com si acabés de *donar la volta al món*, i comentava: (CTILC, 2003)

'as if s/he had just travelled around the world, and commented'

b. I va cridar a la cambrera perquè *en prengués nota.*   (CTILC, 2010)

'And s/he called the waitress so that she would take the order'

This semantic specialization of LVCs is in line with previous research (cf. Italian LVCs in Bratánková, 2013), but it is important to note that this is not general for the corpus sample of Catalan LVCs used in this study. Interchangeability appears to be a multifactorial phenomenon and it is mainly affected by morphosyntax, discourse and semantics.

4.6. Conclusion and final remarks

This chapter has explored the interchangeability of Catalan LVCs and their synthetic counterparts. The present corpus-based study has confirmed findings of previous corpus-based studies on different languages: LVCs with bare nouns have proven to be statistically more likely to be interchangeable, as well as those with unmodified NVEs. The weight on the nominal element, hence, has been found as an important variable when analysing the interchangeability of Catalan LVCs with their synthetic verbal counterpart.

Another relevant variable analysed in this study has been the frequency of LVCs in the Catalan corpus. Although this variable was not included in previous corpus-based studies, its inclusion has been proven relevant, because medium frequency LVCs, followed by high-frequency, favour interchangeability. Hence, LVCs do not seem to be more used when there is no semantic correlation with a synthetic verb, or when they have developed a specific lexicalised meaning that the verb does not have. On the contrary, the most common LVCs are highly interchangeable with a synthetic verb.

The coexistence of Catalan LVCs and their synthetic counterpart can be explained from the identified patterns in the present corpus-based study. In fact, the lack of interchangeability can be explained by referring to different linguistic levels. First, discourse influences the possibilities of LVCs to be interchangeable, especially when the elements of the construction belong to two different clauses (in subordination or

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 4. LVCs and their synthetic counterpart

coordination) and when there is a reference across clauses. Second, some morphosyntactic properties can influence the interchangeability of LVCs and their synthetic verbal counterparts, such as the characteristics of the subject, and the weight of the nominal element in general (determination, number and modification). Finally, the semantic correlation between the LVC and the synthetic counterpart is not always exact, but these cases are a minority.

In future research, data from the interchangeability in the opposite direction (that is, from synthetic verbs into LVCs) should be compared with the findings in this chapter. Only then the existing relationship between LVCs and their synthetic counterparts could be fully explained.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

Chapter 5. Previous approaches to LVCs

This chapter has the goal to provide an overview of previous approaches to LVCs within theoretical linguistics.

First, the status of LVs will be discussed in §5.1, with a special focus on generative grammar. This section will present the following two subsections: lexicalist approaches to LVs (§5.1.1) and syntactic approaches to LVs (§5.1.2).

Second, the most relevant neoconstructionist approaches to LVCs will be discussed in §5.2.

Third, the focus will be set in the analysis each of the LVs under study in this dissertation in §5.3: previous accounts of the LV *give* (§5.3.1), *make* (§5.3.2), and *take* (§5.3.3).

Finally, the chapter concludes with a summary of the main contributions of previous approaches to the analysis of LVCs in §5.4.

5.1. Status of LVs

The status of LVs has been a matter of discussion within theoretical linguistics due to their light meaning while crucially contributing to the meaning of the LVC at the same time. This has been approached from different grammar traditions and their proposals differ in their basic analysis of the LVs and their status.

One of the main differences between lexicalist and syntactic approaches is whether the argument structure is assumed to be listed in the lexicon or generated in the syntax proper (Mendívil-Giró, 2019). Since Jackendoff (1972), lexicalist approaches analyse each lexical item as listed in the lexicon with their argument structure, which is afterwards projected onto the syntax. In Borer's (2005) terms, lexicalist approaches are known as projectionist or endoeskeletal approaches. In contrast, a syntactic or exoskeletal approach has argument structure defined and built by the syntactic principles. That is, it is the position in the structure that imposes the interpretation of the participants that the predicate describes.

This basic difference between both approaches leads to divergent analyses of LVCs. The lexicalist approach concentrates on the complex predication analysis (§5.1.1) while the syntactic approach focuses on the regular verb-object analysis (§5.1.2).

5.1.1. Lexicalist approaches

From a lexicalist perspective, there are different proposals to the analysis of LVCs. In this section, I will first focus on Sanromán Vilas (2011, 2013, 2014, 2017) analysis framed within the meaning–text theory by Mel'čuk (1997); later I will discuss the lexical-functional approach exemplified in Butt (1995, 2010) and Mohanan (2017).

As already introduced in the discussion in Chapter 2, Sanromán Vilas (2017) defends that LVs are not semantically empty and that they are selected by the noun based on its semantics. This corresponds to the hypothesis of semantic compatibility:

> According to this hypothesis, LVs are paradigmatically related to their heavy counterparts and syntagmatically to the nouns within the same LVCs by means of lexical features. At the paradigmatic level, lexical features are the semantic links between the different lexical units existing within the polysemic verb – *tomar*, in this case – where heavy and light senses are included, as well as other collocates. At the syntagmatic level, lexical features are the elements of semantic agreement between the LV and the nouns within the same LVC. (Sanromán Vilas, 2017: 230)

Therefore, her claim is that LVs and their corresponding full verbs have separate lexical entries, although they are evidently related. In fact, Sanromán Vilas (2017: 235) presents the following assumptions as underlying her hypothesis of semantic compatibility:

- For every LV in a language, there is a heavy verb with which it is polysemic: there is the LV *make* and its heavy counterpart *make.*

- Every predicate (or eventive) noun can select a LV with which it shares at least one lexical feature: the noun *promise* selects the LV *make* because it emphasizes the meaning of 'creation'.

- This shared lexical feature(s) agrees also with part of the meaning of the heavy counterpart: the component 'creation' is shared both between the LV and the noun *promise* and the LV *make* and heavy *make.*

Her proposal, also found in Bosque (2004) or Alonso-Ramos (2004), is that there are two lexical units for any verb which can appear in LVCs, but both units are closely related. This approach is useful for the explanatory and combinatorial lexicology since their ultimate goal is to create a catalogue of the lexical units of a language. For this purpose, having two lexical entries is useful to describe the differences of their uses and

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

concentrate on the particularities of both their light and heavy use, however, at the cost of multiplying the number of lexical entries and having no explanatory adequacy.

For a lexical-functional grammar perspective, underspecification is considered the most plausible option. That is, there is a single underspecified representation that can be used as a full verb or a LV depending on the context. In Butt & Lahiri's (2013) study of the diachrony of LVs in South Asian languages, they argue that there is a tight connection between the LVs and their corresponding full verb, which is different from the connection between an auxiliary verb and the verb they are connected with. More specifically, "there is a single underlying lexical entry, which can account not only for the simultaneous synchronic uses of light and main verbs, but also for the historical data" (p. 8).

This goes contra a traditional view of the historical process of semantic bleaching (Hopper & Traugott, 1993) which proposes that LVs are on the cline of grammaticalization. Butt & Lahiri's historical data from Indo-Aryan languages (2013: 27) show that LVs remain stable and are linked to the main verb, because any change affecting the verb has consequences for both the main and the light verb[42]. This is used as proof that they are part of a single underlying entry. In contrast, auxiliaries cease to be linked to the main verb in the historical process and, thus, to be affected.

Apart from the underspecification of the LV, the main claim put forward by these lexical-functional-grammar studies is that the LVC is the result of an argument fusion, where the LV combines with the NVE to yield a single predicate: a complex predicate. In Butt's (2014) words: "Complex predicates are formed when two or more predicational elements enter into a relationship of co-predication. Each predicational element adds arguments to a monoclausal predication. Unlike what happens with control/raising, there are no embedded arguments and no embedded predicates at the level of syntax." (p.7).

In order to test for complex predicates, the diagnostics must be language specific. The main question is whether the elements involved in the complex predicates have

---

[42] These findings are also in line with Elenbaas' (2013) study on synchronic and diachronic data of English LVCs which do not display signs of grammaticalization and are, thus, considered synchronic variants of full verbs from early stages. For instance, in (i), there is an example from the *Works of Sir Thomas Malory* from the 15th Century, a moment when LVCs already co-existed with their synthetic counterpart.

> (i) But sitthyn I have *made a promyse,* with that thou wolte promyse me to beare upon the a shylde …   (Elenbaas, 2013: 53)
> 'But since I have made a promise, with that you will promise me to bear a shield upon you'

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

combined to form a single co-predicational domain. In Romance, they famously include clitic climbing, as in the French causative construction (Alsina, 1992; Butt, 2010), in (1). In this case, this construction is considered a LVC with the causative LV *faire* followed by the infinitive verb *partir* constituting a monoclausal predication[43].

(1) a. Jean a fait partir Marie.

Jean has made go Marie

'Jean made Marie go.'   (Rosen, 1989: 22)

b. Jean l'a fait partir.

Jean her has made go

'Jean made her go.'   (Rosen, 1989: 23)

The rest of languages from which tests are provided to test monoclausality in LVCs belong to the Indo-Aryan language family (i.e. Hindi and Urdu in Butt, 1995) and Asian languages (i.e. Korean in Choi, 2005), where two verbal elements are involved in the described LVCs. These structures are distinct in nature from the verb-noun constructions which constitute the LVCs dataset in the four languages under study in this dissertation and will not be further discussed here.

Mohanan (2017: 2) builds on the previous traditional conception of LVCs as complex predicates (Butt 1995, 2010) and proposes that such a complex predicate "is composed of two morphs, both semantically predicative, jointly determining the argument structure of a single syntactic clause". Therefore, the fact that the LVs and the NVE are independent semantic predicates conforming a single syntactic clause distinguishes LVCs from noun incorporation.

Regarding the status of the LV, both Butt (2010) and Mohanan (2017) argue that it is a grammatical element, that is, a semi-lexical verb which constitutes a word-level unit functioning as a member of a complex predicate. In fact, LVs and their non-light use as full verbs have a single underspecified representation used as full or LV depending on the context.

When a LV followed by a NVE is used in monoclausality tests, such as the clitic climbing test in (1), it is not a valid test because the structure is that of a verb and an object, as in (2) with the Catalan LV *fer* ('do/make') and the NVE *un regal* ('a gift').

---

[43] See Alsina (1992) for an analysis of causatives in Romance as complex predicates.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

(2) a. En Víctor ha fet un regal a la Griselda.

'Víctor has given Griselda a gift'

b. En Víctor li ha fet un regal.

'Víctor has given her a gift'

Other tests for monoclausality in Romance languages are found in Jardón Pérez (2022: 43-46) but they are restricted to V-V constructions. For instance, a prototypical monoclausal structure is the perfect tense in Spanish, where the auxiliary verb does not show any form of agreement with the object, as in (3).

(3) He escrito diez cartes

Haber-PRS.1P.SG write-PART ten letters

'I have written ten letters'

Two tests can be applied to the perfect tense to proof its monoclausality: first, strict adjacency is obligatory in monoclausal structures, as in (4); second, the possibility to form a *how* question is disallowed, as in (5).

(4) *He las cartas escrito

Haber- PRS.1P.SG the letters written- PART

(5) *¿Cómo has las cartas?

How haber- PRS.2P.SG the letters

However, the application of these two tests for monoclausality in a verb-noun LVC of the sort analysed in this dissertation shows opposite results: in (6), the NVE can move to other positions of the sentence, and it does not need to be strictly adjacent to the LV; and (7) allows a *how* question – although only when the NVE is included in the question through pronominalization.

(6) a. Un regal li ha fet el Víctor a la Griselda.

A gift her.DAT has.AUX made.PART the Víctor to the Griselda

b. N'ha fet un, de regal.

PARTV.has.AUX made.PART one, of gift.

(7) Com *(ho) ha fet en Víctor?

How it has.AUX made.PART the Víctor

The fact that the sample comprising this dissertation is based on verb-noun constructions differs in nature from that analysed as complex predicates in previous studies, which focus on complex verbal predicates (with two verbs, or LV and co-verb).

It also results on a lack of monoclausality in the eventive LVCs analysed in this dissertation, as shown in (2), (4) and (5).

5.1.2. Syntactic approaches

LVCs have also been analysed in the syntax-semantics interface study of the language. In this section, I will first discuss Bruening's (2016) proposal of LVs as regular verbs in English; then I will briefly comment on Ramchand's (2013) analysis of the structural and conceptual meaning of LVs (as already discussed in Chapter 2, §2.1.1.); and I will finally present Borer's (2005) model on structuring sense.

In opposition to Butt's claim (presented in §5.1.1), Bruening (2016) argues that LVs in English are just regular verbs in verb-noun LVCs, and there is no need for a complex predication analysis. Instead, he argues that LVs are just regular verbs with little semantic specification and their complements are just regular complements, which can also be found in constructions with full verbs, as in (8).

(8) a. Take a careful look at this! (light verb)

b. I recommend a careful look at this. (full verb)

In the same way as LVs are identical in form to full verbs, the nominal element is also a regular noun phrase with the characteristic of being eventive. In this case, the event denoted is a separate event from that of the LV (9a), and it does not need a LV to be eventive, since it is eventive on its own (9b).

(9) a. She had [three quick espressos] and then left to the coffee shop.
b. [Two short side trips] is all we have time for.   (Bruening, 2016: 55)

For Bruening (2016), there is no need for a special grammatical operation just for light verbs. In his approach, only two aspects are necessary to account for LVCs: i) the eventivity of the noun, and ii) the need for control into the noun phrase.

The need for control is explained by the fact that the complement of the LV plays a role in the argument structure of the LVC: "its own arguments are controlled by the argument structure of the verb. This control is something we see outside of light verb constructions and is not something that we need to build a theory of just for light verbs" (Bruening, 2016: 58). In particular, the logical arguments of the NVE can be controlled by arguments of the LV. As in (10), LVCs with *give* allow an indirect object even when

the NVE does not have a logical argument (contra Huddleston & Pullum, 2002: 293). At the same time, in (11) the LVCs with NVE that usually take a logical argument can appear without the IO.

(10)   a. She turned and *gave the audience a deep sigh*.

        b. Come on, *give us a vomit*.

        c. The baby *gave us a giggle*.

(11)   a. She *gave one violent kick* and broke free.

        b. She *gives massages* all day long.

        c. Tink the horse *gives kisses* for treats.   (Bruening, 2016: 55)

Bruening's proposal is to build a theory of control into nominals, and shows that, in the case of *give*-LVCs, there are two overlapping control relations, as in (12).

(12)   She$_i$ gave him$_j$ [a KICKER$_i$ kick KICKEE$_j$ in the teeth]

                      (Bruening, 2016: 58)

His suggestion is thus to modify the Agree Theory of Control developed in Landau (2008) by relativizing Agree, while allowing these two overlapping control relations. However, existing theories of control do not permit overlapping, as in (12).

One problem for this approach is already pointed out by Bruening himself regarding LVCs with a causative interpretation, where the indirect object is not an internal argument of the noun, but rather its external logical argument, as in (15). This kind of constructions should need a different analysis, which cannot be framed under the control theory.

(13)   That *gave us a laugh*. (laugher = us)

In the study of control into nominals in Romance languages (Alba-Salas, 2006), LVs are considered to induce obligatory control in their nominal complement, as in (14).

(14)   La Mònica$_i$ (li) farà [PRO$_i$ una trucada (*del Pere) a l'Eva]

        The.FEM Mònica$_i$ (to him) make.3SG.FUT [PRO$_i$ a call (*of.the Pere) to.the Eva]

        'Mònica will make the call (*of Pere) to Eva.'

However, according to Landau (2013, 2015), both in Romance and Germanic LVCs there is no PRO, since they do not project a structural subject. In (15a), the promiser (*She*) is linked to Sandy thanks to the obligatory control exerted by the light verb (*made*). In (15b), the PRO is carried over inside the promise-DP.

(15)   a. She made Sandy a promise.

        b. the promise that [[the candidate]$_i$ made] [PRO$_i$ to lower taxes]

(Landau, 2015: 129)

In recent syntactic approaches, there is no need for control to account for the argument structure in LVCs. The main tendency is to consider that the LV is inserted into the functional layer *v*, which enables its contribution to the event and argument structure of the whole construction.

According to Ramchand (2013), the argument and event structures of the LVs are identical to those of the full verb (that is, their use in non-light contexts)[44]. The main difference between the light and full uses of the verb *give* lies in the nature of the complement, that is, the NVE: if the complement is eventive, the interpretation of the construction is that of an LVC. The focus to LVCs is shifted from the LV itself to the nature of the NVE, and of the construction as a whole.

From a purely syntactic approach, Borer (2005) proposes that the arguments of lexical items are not specified in their lexical entry. Instead, they are interpreted according to where they are in the syntactic structure, specifically in the eventive structure. "The proposed dividing line, I believe, is a crucial one, and distinguishes between what is grammatically real – structures and formal properties of functional items – and what may be very real, but not grammatically so –properties of substantive vocabulary" (2005: 8). That is, the different elements will be interpreted depending on their place in the syntactic structure, where the syntactic computation is carried out by functional items which merge with listemes[45], which are devoid of grammatical structure (2005: 17). Thus, Borer's model does not support that verbal polysemy can be explained by the existence of several lexical entries, but rather proposes that polysemy takes place in the syntactic structure. In

---

[44] The contribution of the LV to the whole construction according to Ramchand has already been discussed in more depth in Chapter 2 (§2.1.2).

[45] For Borer (2005), *listemes* do not have any internal grammatical structure: they have no category, no morphological marking, and no argument structure either.

fact, she postulates that the arguments are related to the event over the verb itself, which is "a modifier of that event, rather than a determinant of its interpretation" (p. 46).

From Borer's perspective, listemes can be inserted in different functional structures. This can explain why there are new uses of different verbs (as proposed in Galbarini, 2017: 15). For instance, the Spanish verb *dar* ('give') is inserted in a transitive telic structure in ditransitive sentences in LVCs such as *dar comida* ('give food, to feed') in (16).

(16)   Juan le dio comida al perro

        'Juan gave food to the dog'

In a similar line, Cuervo (2010: 2) analysed the LV *dar* 'give' in Spanish within this perspective arguing that it is the argument structure of the whole construction which determines the verbal meaning and the interpretation of the arguments. Cuervo proposes different constructions for LVCs like *dar hambre/sueño/náuseas* (lit. 'give hunger/sleepiness/nausea'), which parallel psych verbs like *gustar* ('like'), and ditransitive LVCs like *dar apoyo/ánimo* (lit. 'give support/encouragement'). Both structures are different due to the applicative projection: the former structure bears a high applicative with a vP as a complement (17), whereas the latter has a low applicative with a DP as a complement (18).

(17)   Al técnico le dan rabia las protestas

        The coach.DAT CL.DAT give.PL fury the complaints.NOM

        'The complaints make the coach furious'
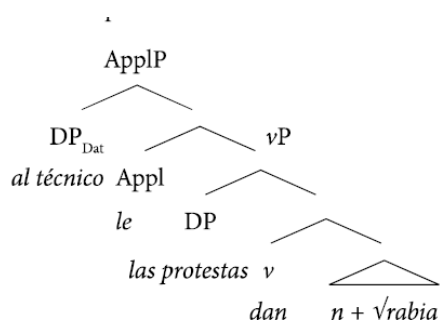
                                                        (Cuervo, 2010: 146)

(18)   El plantel siempre le da apoyo al técnico

        The team.NOM always CL.DAT gives support the coach.DAT

        'The team always supports the coach'
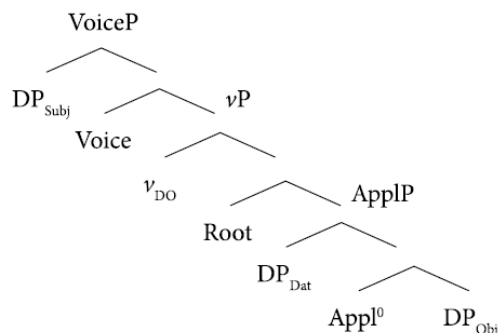
(Cuervo, 2010: 145)

This dissertation proposes an analysis of LVCs within a neocontructionist approach to argument structure which defends that the syntax in the vP determines the event structure of the predicate, and the thematic interpretation of its arguments. I will develop this approach and its relevance for the LVCs analysis in the next subsection.

5.2. Previous analyses of LVCs from a neoconstructionist approach

In the neoconstructionist approach, argument structure is a syntactic configuration built from a limited set of possible combinations from two sets: relational elements (functional heads which project structure, such as v head, Path or Place) and non-relational elements which are unable to project structure and appear at Compl or Spec position (such as roots or DPs) (Mateu, 2002; Acedo-Matellán, 2016). Within this framework, argument structure is syntactic, and it is to be identified with the syntactic structures projected by lexical heads.

The basis for this framework is the work by Hale & Keyser (1993: 55) who propose an analysis of unergative verbs as underlyingly transitive, based on unergative activity verbs in Basque and other languages, which are built as a sort of LVC, as in (19). That is, unergative verbs in Basque are formed out of a nominal N and a light verb V.

(19)   a. barre egin ('laugh do') = to laugh

   b. lo egin ('sleep do') = to sleep

Their proposal is that both LVCs and unergative verbs share the same structure with the difference that a synthetic unergative verb involves the incorporation of the nominal head N into an abstract V (20) unlike LVCs, where there is no incorporation needed because the complement stays in its base position (21).

(20)
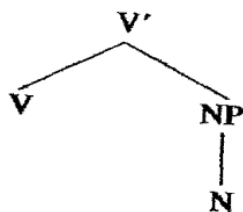
(Hale & Keyser, 1993: 54)

(21)
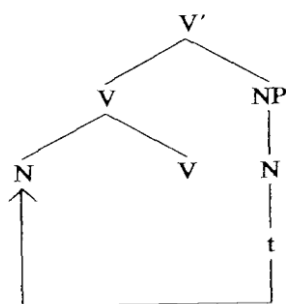
(Hale & Keyser, 1993: 55)

Based on Hale & Keyser and other elements in Distributed Morphology such as Late Insertion, Mateu (2002), Mateu & Acedo-Matellán (2012) and Acedo-Matellán (2010, 2016) argue that unergatives are activity events where the root incorporates into the v head: the interpretation of the unergative *dance* is not exactly the same as *do a dance*. The difference in interpretation is based in the event structure and it stems from the nature of the objects, when there is a DP (*a dance*) the object carries functional structure which is to be semantically interpreted whereas the root DANCE does not have an object interpretation (Acedo-Matellán, 2014).

Argument structure configurations build different types of vP (Acedo-Matellán, 2016: 35): the unergative *danced* is represented in (22a), where the complement of v is a root (DANCE), while the creation event *did a dance* is the configuration in (22b), where the complement is a DP (a dance), and it behaves like a canonical argument.

(22) a. Unergative *danced*   b. Transitive creation *did a dance*
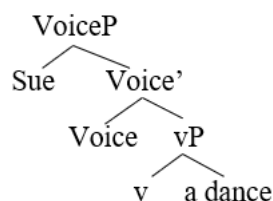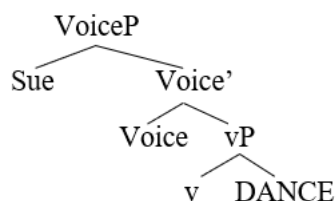
(Acedo-Matellán, 2016: 31)

From this perspective, LVs contribute the event structure in the LVC from their position in v, but no conceptual semantics because they lack a root. Instead of assuming that there are two verbs *do, have, give, make*, the assumption is that these verbs have an underspecified semantics and combine with the complement in different positions in the syntactic structure.

Even after the incorporation of the root in unergative verbs (22a), the v head can take a DP complement which is an incremental theme which provides a scale to the event predicated by the verb and has a creation consumption reading (Mateu & Acedo-Matellán, 2012). This complement can either be a hyponomous object (23a) or even a cognate object (23b).

(23) a. John dances *a tango*.

b. John dances *a beautiful dance*.

While these objects in (25) have received attention in the literature (i.e. Macfarland, 1995; Massam, 1990; Pereltsvaig, 1999, 2002; Real Puigdollers 2009, 2013; Fullana et al., 2022), their structural position is still under debate.

In contrast, LVCs do not generally accept a second DP because the NVE occupies the object position and there is no incorporation involved (24), with the exception of *dar vuelta* 'give turn' in Argentinian Spanish (25) which accepts a direct object that can be pronominalized (25b), as pointed out in Galbarini (2017).

(24) a. *Dio respuesta un mensaje.

Gave.3SG answer a message

b. *Hicimos análisis un estudio.

Made.1SG analysis a study

(25) a. De la mano de Messi, Barcelona dio vuelta *el partido.*

Of the hand of Messi, Barcelona gave.3SG turn the game

'Thanks to Messi, Barcelona turned the game around'

b. Con uno menos, Estudiantes se *lo* dio vuelta a River.   (Google)

With one less, Estudiantes SE it.AC gave.3SG turn to River

'With one (player) less, Estudiantes turned it (=the game) around for River'

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

    c. Lo puso al revés cuando dio vuelta *el pizarrón*.   (Google)

    It.ACC put in.the backwards when gave.3SG turn the blackboard

    'He/she put it backwards when turning the blackboard'

According to Jiménez Martínez (2016: 94), this was also a possibility in Latin, although not very frequent. Some LVCs which are very fossilised are capable to take a second object as in *mentem habere* 'have in mind' or *gratias agere* 'give thanks'.

As far as the other languages under study in this dissertation are concerned, there are no studies investigating LVCs allowing an extra direct object besides the studies for Spanish (Galbarini, 2017; Matute Martínez, 2012). Therefore, the basic assumption is that generally the NVE in LVCs occupies the position of the complement and no incorporation process takes place, except for the case of *dar vuelta* in (25), as further discussed in Chapter 6.

Still within a neo-constructionist approach to argument structure, in line with Acedo-Matellán's (2014) proposal for *do*, Myler (2016) studies possession with *have* focusing on English and Potwari. Myler's proposal is that the LV *have* is a realization of *v* and it contributes an eventuality variable, whereas the NVE provides the content of the event. In (26), the verb *have* takes a DP as a complement which denotes an event or state, while Voice introduces the external argument in Kratzerian terms (Kratzer, 1996).

(26)   We had a conversation
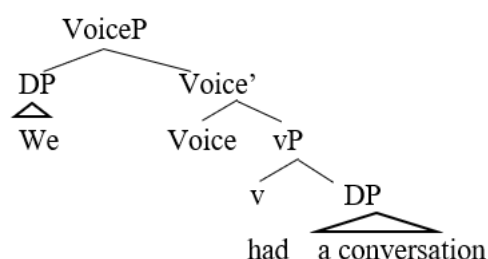
                             (Adapted from Myler, 2016: 278)

One particularity of LVCs with *have* noticed by Myler (2016) is that the NVE must be a simplex event nominal (à la Grimshaw, 1990), as in (27), and complex event nominals are rejected, as in (28).

(27)   a. We had a conversation.

        b. We had an argument.

c. I had a bath.

d. You'll have trouble getting that to work.

(28) a. * The Romans had {a/the} destruction of the city.

a'. * The city had {a/the} destruction.

b. * John had {a/the} payment of his bills.

b'. * His bills had {a/the} payment.

This is due to a particular characteristic of the verb *have* (not only in its light use) which is related to delayed gratification: "That is, complex event nominals systematically resist delayed gratification of the sort that I have argued takes place in relational have sentences. This is true regardless of whether the agent or the theme argument of the derived nominal ends up as the subject of have" (p. 278).

Following Alexiadou & Grimshaw's (2008) account, complex event nominals have verbal substructure within the nominal domain (29a), while simplex event nominals (29b) and relational nouns (29c) do not. It is precisely the absence of v in the substructure what would account for their compatibility with delayed gratification and, thus, with LV *have*.

(29) a. [[√ destruct−$\emptyset_v$]vP -ion_n]nP

b. [√ argue -$ment_n$]nP

c. [√sister -$\emptyset_n$]nP

This property of the LV *have* is not shared by other frequent LVs such as *give, take* and *make,* which implies that different LVs must be realizations of v in different structural environments (Myler, 2016: 419). The question that arises is what these environments are and how they can be described, taking into account that there are differences in closely related languages or within the same language in the choice of an LV with a specific NVE.

5.3. Previous accounts of the LVs under study

From a semantic perspective, Colominas (2001) analyses the three verbs (*fer, prendre, donar*, Catalan for *make, take* and *give*) as dynamic verbs, which share the agentivity of the event that they select. In the analysis of dynamics verbs, there are some aspectual properties that set them apart. While *fer* ('make') is related to accomplishments (30a), the verbs *prendre* ('take') and *give* ('donar') denote achievements (30b,c).

261

(30)　　a. *fer una cosa en una hora*

　　　　'make a thing in an hour'

　　　　b. **prendre una cosa en una hora*

　　　　'take a thing in an hour'

　　　　c. **donar una cosa en una hora*　　　　　　(Colominas, 2001: 203)

　　　　'give a thing in an hour'

In terms of argument structure, there are some basic differences regarding the arguments involved. While *fer* ('make') involves two arguments (agent and theme), the other two verbs involve three arguments (agent, theme and experiencer / beneficiary), as seen in (31). From a general perspective, the main difference is that the first is a verb of creation while the other two are verbs of transfer.

(31)　　a. *fer* (e1, agent, tema) <α existir (e2, tema)

　　　　Make (e1, agent, theme) <α exist (e2, theme)

　　　　b. *prendre* (e1, agent, tema) <α tenir (e2, tema, destinatari)

　　　　Take (e1, agent, theme) <α have (e2, theme, receiver)

　　　　c. *donar* (e1, agent, tema) <α tenir (e2, tema, experimentador)

　　　　Give (e1, agent, theme) <α have (e2, theme, experiencer)

　　　　　　　　　　　　　　　　　(Colominas, 2001: 204)

Moreover, the main difference between the two transfer verbs lies on the nature of this third argument: in *prendre* ('take'), the event is directed towards the same participant undergoing the action (32a); while in *donar* ('give') the beneficiary is necessarily a third argument (32b). Thus, it is the argument structure of these two verbs which allows them to select different kinds of nominalizations although they are both verbs of transfer.

(32)　　a. prendre (e1 X, Y)

　　　　b. donar (e1, X, Y, Z)

In Colominas' analysis based on the proposal by the Head-driven Phrase Structure Grammar (based on Pollard & Sag, 1987) and The Generative Lexicon by Pustejovsky (1995), she coincides with previous studies (as discussed in §5.1) that there is only one lexical entry for the verb which is underspecified in the use of the verb in LVC. Thus, the LV preserves the properties of its use as a full verb. Moreover, in her analysis of Catalan LVCs, Colominas (2001) concludes that the semantics resulting of LVCs such as *fer una*

traducció ('make a translation') or *donar una resposta* ('give an answer') is similar to that of the synthetic counterpart *traduir* ('translate') and *respondre* ('answer'), with an exception: the fact that the LVC can make reference to the nominal through the modification, as with *excel·lent* ('excellent') in (33).

(33) En Joan ha fet una anàlisi de les dades excel·lent en un temps record.

'Joan has made an excellent analysis of the data in record time'

In sum, this proposal concentrates on the semantics representation of LVCs with special emphasis on the compatibility between the NVE and the LV, but it does not focus on the internal structure within the NVE and how it can affect the structure of the whole construction.

Taking the differences detected in the three Catalan LVs by Colominas (2001), in the following subsections, previous syntactic accounts of the three verbs will be summarized and compared: *give* (§5.3.1), *make* (§5.3.2) and *take* (§5.3.3).

5.3.1. Previous accounts of the LV *give*

The LV *give* involves a ditransitive construction, and it has caught the attention of several scholars from different perspectives. Ditransitives and double object constructions have been widely studied both in Germanic (e.g. Harley, 2002; Pylkkänen, 2008) and Romance languages (e.g. Cuervo, 2003; Pineda, 2015).

Within Distributed Morphology and Marantz's (2001) proposal of little categories, Kornfeld (2004: 220) proposes an analysis for the Spanish LV *dar* in two contexts: with bare nouns, like *dar clase* 'give lesson', and determined nouns, like *dar una clase* 'give a lesson'. For the former, the analysis in (34a) is a complex predicate where the root is merged directly with the *nP*, which are then taken as a complement of little v. For the latter, the DP is introduced in an intermediate projection ROOT' (34b).

(34)
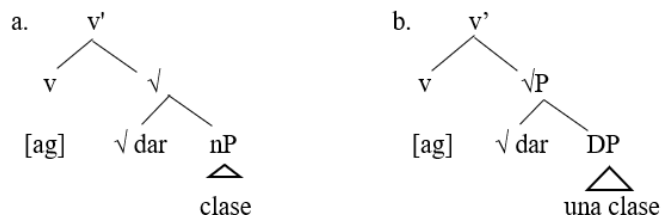
(Adapted from Kornfeld, 2004: 220)

This proposal argues that it is not true that in complex predicates there are no functional categories. However, the determiner in the NVE (DP) does not allow it to be merged directly with the root (34a).

In both analyses (34) the LV is a lexical root, unlike in Acedo-Matellán (2014) for *do a dance* in (22) and Myler (2016) for *have a conversation* in (26). However, Kornfeld (2004) argues that the option of considering the verb as a direct instantiation of a little *v*, and not a root, could also be plausible since the LV in such constructions is ambiguous regarding its functional or lexical nature.

On the one hand, the LV *dar* in some contexts such as *dar alimento* ('give food') can be considered to correspond literally to the lexical meaning of *dar* ('give'), as in *dar un libro* ('give a book'). Hence, the root must be identified in both cases. From the little categories in Distributed Morphology (Marantz, 2001), the root in LVs seems to be superfluous and an alternative analysis can be put forward, as in (35).

(35)

(Adapted from Kornfeld, 2004: 222)

264

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs
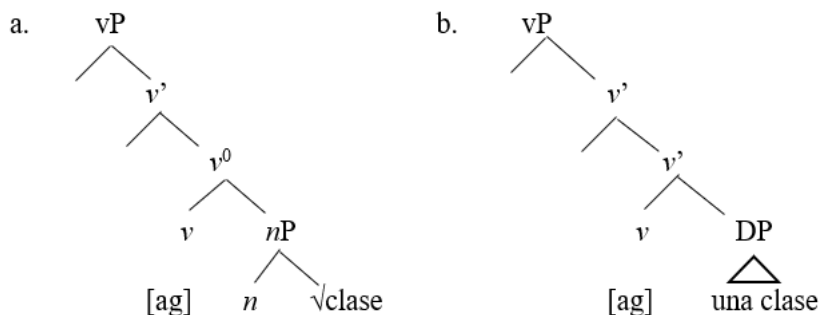
Kornfeld (2004) prefers the analysis represented in (35) because it allows the differentiation of Spanish with languages allowing object incorporation with lexical verbs, as in Mithun's (1984) typology. That is, object incorporation in Spanish is only possible with LVs, as in (35a), while languages like Mapuche allowing incorporation with lexical verbs (as in *nufa -mar 'ra-ley*, lit. 'hunt(ing)-hare-is') should be analysed as in (34a) where there is a merge of the *nP* with the verbal root.

The analysis of the LV as incorporated in the functional layer *v* is also the most common within the neoconstructionist approach to argument structure after Kornfeld (as in Acedo-Matellán, 2014; Myler, 2016).

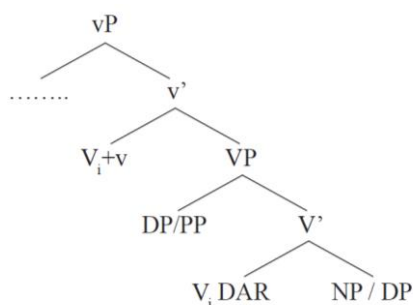At the same time, the distinction between complex predicates (as in *dar clase*) and expressions with a DP (as in *dar una clase*) lies in the kind of phrases it combines with: an *n*P or a DP, and the level of introduction within the tree (v^0 vs. *v'*). For Kornfeld (2004), the lack of D in the first case is enough to trigger the complex predicate analysis, since only DPs can behave as arguments of a phrase and, thus, the bare nominal (*n*P) has to merge at v^0 in order to be legitimized as a constituent. This distinction between bare nominals and DPs will be crucial for the analysis of the dataset in this dissertation, as presented later in Chapter 6.

Another proposal on ditransitive LVCs is that of Choupina & Brito (2018) for Portuguese *dar* ('give'). They distinguish between their use in ditransitive constructions, as in *dar um espurro*, and in transitive constructions without a beneficiary, as in *dar um paseo*. In (36), the indirect object is generated as Specifier of VP (following Brito, 2010; Gonçalves, 2016), and the LV occupies the V position, identical to its use as a full verb, but it builds a complex predicate with the deverbal NVE. However, in (37) the LV *dar* is generated directly in v and it also builds a complex predicate with the NVE, but the structure does not have any layer to introduce a beneficiary object.

(36)

(37)

```
                    vP
                 ╱      ╲
          ……..            v'
                       ╱      ╲
                 V leve        VP
                             ╱    ╲
                           V       NP/DP
```
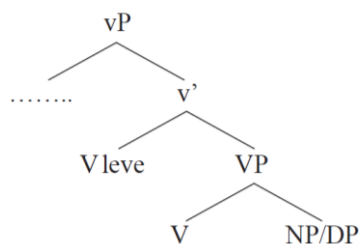
(Choupina & Brito, 2018: 172)

For English *give*, Bruening (2015) puts forward an analysis with three variants: causative *give,* produce *give* and true light verb *give,* as in (38).

(38)

```
                    VoiceP
                 ╱        ╲
            NP              Voice
            ⌒            ╱        ╲
           she       Voice         ApplP
                              ╱          ╲
                           NP              ApplP
                           ⌒           ╱         ╲
                          him        NP            Appl
                                     ⌒          ╱       ╲
                                   a kick   Appl_for      VP
                                                        ╱    ╲
                                                      V       NP
                                                      g-      ⌒
                                                            a kick
```
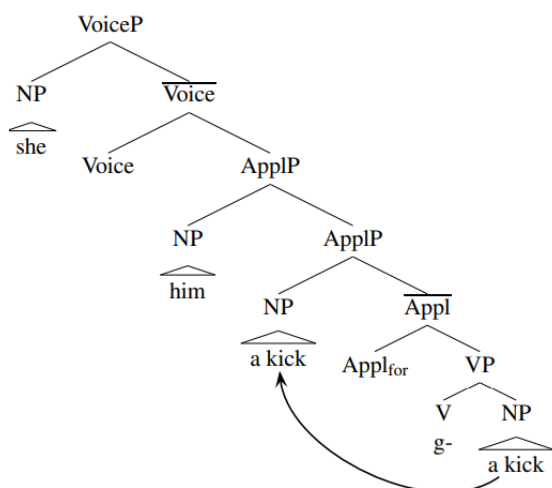
(Bruening, 2015: 24)

The main difference between the last two is that only true LV *give* allows a depictive secondary predicate. In his proposal, this particularity can be explained because the true LV combines with its complement by conjunction, not by Application as in (38), and this can allow a depictive to modify the indirect object. According to Bruening (2015: 25), "this proposal for true light verb *give* explains why its indirect object can be modified by a depictive secondary predicate, while other indirect objects cannot be modified. The difference is the mode of combination: the first object of a true light V combines via predicate modification rather than by function application". This analysis could also be extended to other transitive LVs, as *take* and *have*, taking into account that the shared argument might be different depending on the presence/lack of Voice: in *take* the shared argument is the external argument of the noun, while in *have* it is the internal argument.

The fact that LVCs allow the selection of an optional IO (both with *give* and *make*) needs to allow an applicative head within their structure: "applicatives are now widely considered to include any element introducing a non-core argument, whether or not this element surfaces as an overt applicative morpheme" (McGinnis, 2008: 1227).

Pylkkänen (2008) distinguishes between two classes of applicatives: high and low applicatives. The former are generated outside the VP and they relate the VP with an event-related applied argument (i.e. beneficiary) (39a). The latter are generated within the VP and they relate the theme argument and the verb to an applied argument (i.e. recipient or source) (39b). For Pylkkänen (2008), low applicatives involve a dynamic transfer of possession of the theme to or from the applied argument, and she applied this analysis to English double object constructions.

(39)

a. High applicative
b. Low applicative
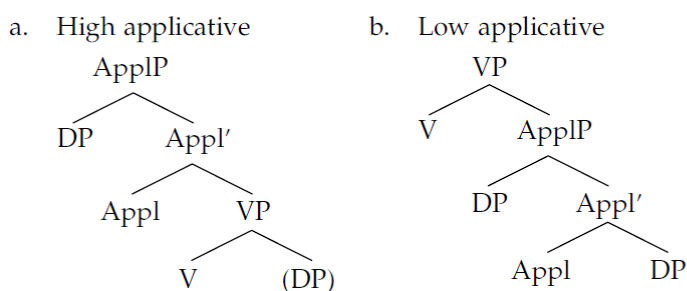
(McGinnis, 2008: 1227)

Cuervo (2003) applies a similar analysis with a low static possessor applicative to Spanish ditransitive constructions with a clitic-doubled dative DPs, as in (40).

(40)   Pablo le besó la frente a Valeria.

Pablo CL.DAT kissed the forehead.ACC Valeria.DAT

'Pablo kissed Valeria on the forehead.'

This analysis is in line with previous research (Masullo, 1992; Demonte, 1995; Romero, 1997) which draws a parallel between Spanish ditransitive constructions with dative clitic doubling and English double object constructions, as well as non-doubled ditransitive constructions corresponding to prepositional constructions with *to*-datives in English. For Cuervo (2003), *dar*-LVCs (i.e., *dar apoyo* 'give help', *dar permiso* 'give permission', *dar aliento* 'give breath', *dar muerte* 'give death') with clitic doubling are

analysed with a low applicative (as in (20) above). The dative case is expressed in the *a*-DPDat while the clitic is introduced in the Appl.

Folli & Harley (2013) also include a low applicative in the *dare*-LVCs with –*ata* nominalizations in Italian, as well as a v$_{CAUSE}$ which allows the introduction of the external argument, as in (41). The Applicative head relates a second argument to the event nominal, expressing an affectedness relation between the applied argument and the NVE.

(41)   Maria ha dato una sgridata a Gianni

'Maria has given Gianni a reprimand'
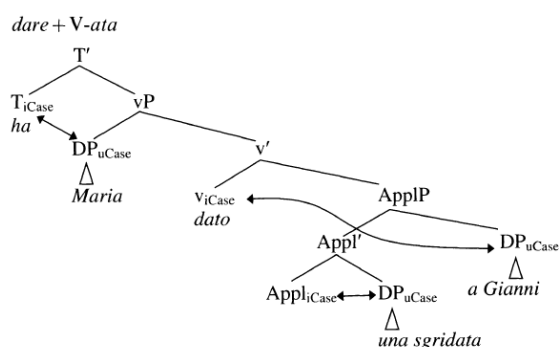
(Folli & Harley, 2013: 95)

The underlying v$_{CAUSE}$ is identical in LVCs with *dare* and *fare* (as discussed in upcoming subsections) with the difference that *dare* is the spell-out of v-Appl while *fare* is the spell-out of pure causative *v*, which is incompatible with a dative argument (Folli & Harley, 2013: 102).

For other Romance languages, Acedo-Matellán & Pineda (2019: 187-188) argue that the dative in LVCs is an added argument through an applicative (following Pylkkänen, 2008), which can explain some correlations between the use of *dar* ('give') in Spanish and *fer* ('do') in Catalan in combination with the same nominal (42). Under this hypothesis, both LVs can combine with the applicative, which is a dative-introducing head.

(42)   Sp. dar un beso ~ Cat. fer un petó 'give/ make a kiss', Sp. dar un abrazo
~ Cat. fer una abraçada 'give/make a hug'

With this model, there is no need to distinguish between LVCs with *give/geben/dar/donar* in ditransitive versus transitive constructions. Both structures share the same transitive base, which can select an Applicative head in the right contexts where an extra argument needs to be introduced. Pineda (2014, 2015) proposes a different Appl

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

analysis from that of Cuervo (2003, 2008), but still advocates for an Applicative analysis for the introduction of a third argument in Romance languages. The form of this phrase and the specific properties it bears will be further discussed in Chapter 6.

To sum up, there is consensus that the LV *give* (and its corresponding variants in the different languages) would be introduced in the functional event layer v, and that the argument structure needs an applicative to introduce the IO, as any other ditransitive and double object constructions. It is also highlighted that the DO is treated differently, depending on whether it is a DP or a bare nominal. The amount of structure within the nominal is key for its analysis, as it will be further discussed in Chapter 6.

5.3.2. Previous accounts of the LV *make*

The LV *make* (and its corresponding LVs in the languages under study: *machen*, *fer*, *hacer*) is the prototypical creation verb. For Folli & Harley (2013), LVs are always introduced in the v position (as seen before for the LV *dare*). In the case of the Italian LV *fare,* the flavour is v_{cause} which allows the introduction of an external argument in their framework, (43).

(43)   Maria ha fatto una risata
        'Maria has made a laugh'
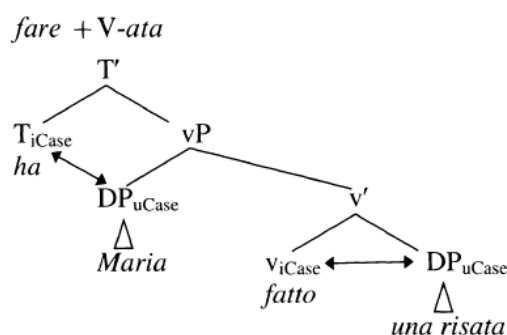
(Folli & Harley, 2013: 95)

Sanromán Vilas (2014) analyses the Spanish verb of creation, *hacer*, as having two arguments (or actants in her theoretical framework): the external argument (X) and the internal argument (Y), as in (44).

(44)   X makes Y = X causes that Y exists

In this schema, the LV *hacer* as a creation verb can also be included in ditransitive constructions with a third argument: in that case, the indirect object is a beneficiary which is conceptualized as a receiver although the transfer is not literal (Ortiz de Ciscomani,

269

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

2006). As discussed in the previous subsection, the introduction of the third argument is through an applicative.

A similar proposal is put forward for Portuguese by Gonçalves et al. (2010), who argues that the verb *fazer* ('make'), in its full use as a lexical verb, is a causative predicate which implies a change of state, and the benefactive is an optional argument which can be appear as dative or oblique prepositional phrase (45a). However, the LV *fazer* can only cooccur with a dative (45b).

(45)   a. O Pedro fez uma casa na árvore (às / para as crianças).

        The Pedro made.3SG a house on.the tree (to / for the children)

        'Pedro made a tree house (to / for the children)'

        b. O filme fez muita aflição / impressão às crianças.

        The film made. 3SG a lot anguish / impression to.the children

        'The film anguished / impressed the children'

This is also applicable to Spanish and Catalan instances with the verbs *hacer* (46) and *fer* (47).

(46)   a. El carpintero hace una mesa a/para los clientes.

        'The carpenter makes a table to/for the clients'

        b. El masajista hacer un masaje a/*para el cliente.

        'The massage therapist makes a massage to/*for the client'

(47)   a. En Pere va fer una casa de fusta al/pel gos.

        'Pere made a wood house to/for the dog'

        b. En Pere va fer una impressió a/*per les criatures.

        'Pere made an impression to/*for the children'

        c. En Pere va fer un petó a/*per les criatures.

        'Pere gave a kiss to/*for the children'

In languages where LV *make* and *give* can coexist with the same nominals, such as Spanish, the LV *make* only selects NVE which are created, whereas the LV *give* can occur with nominals which are pre-existent to the moment of utterance. In both verbs, there can also be a transfer which would enable the presence of a beneficiary/receiver argument (De Miguel, 2011; Sanromán Vilas, 2014). However, the LV *make* emphasizes the creation of the event, while the LV *give* highlights the transfer.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

In sum, the LV *make* (and its corresponding verb in all languages) is a creation verb, even when a third argument (IO/Beneficiary) is introduced in the construction. Its basic argument structure can explain why it is considered one of the most basic LVs in several languages, such as in Catalan, and why it coexists with other LVs in Catalan (*fer/donar una ullada* 'make/give a look'), Spanish (*hacer/dar masajes* 'make/give massages', *declaraciones* 'statements', etc.) and English (*give/make a speech*). However, to my knowledge, the German *machen* 'make' has not been studied in detail, and especially not from a syntactic account[46].
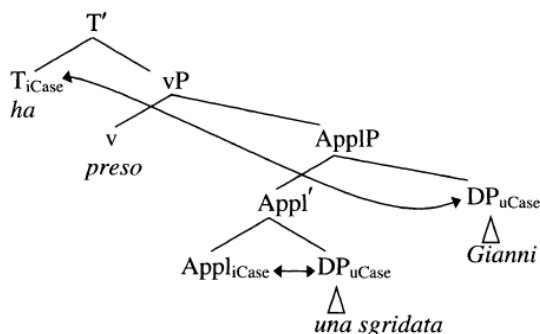
5.3.3. Previous accounts of the LV *take*

The LV *take* (and its corresponding LVs in the languages under study: *nehmen, prendre, tomar*) is not in complementary distribution with the other two LVs analysed: *give* and *make*. While the other two LVs can select the same nouns inter- and intra-linguistically (48), the LV *take* has some properties which contrast with the other two in the languages under study, although it does coexist with *make* (49a) and *have* in different English varieties (49b).

(48)   a. give / make a speech

 b. fer / donar una ullada (Catalan, lit. 'give/ make a look')

 c. dar / hacer un masaje (Spanish, lit. 'give/ make a massage')

(49)   a. take a decision (British English) vs. make a decision (American & British English)

 b. take a bath (American English) vs. have a bath (British English)

Folli & Harley (2013: 94) state that the Italian LV *prendre* is in a paradigmatic relation with *dare* and *fare*. Their contrast is based on the type of v element involved in the syntactic context, as in (50). The position of the LV is v in all three cases but there are variations: the flavour of v in *take* is v_{BECOME,} whereas in the LVs *make* and *give* is v_{CAUSE}; and the applicative projection, which introduces the internal argument.

(50)   Gianni ha preso una sgridata.

 'Gianni has got a reprimand'

---

[46] Schmid (2016) compares the use of Cat. *fer* and Ger. *machen*, the latter is more colloquial and has less combinatory possibilities than the Cat. *fer* 'make'. A more restricted use in German could have influenced its lack of attention in the theoretical linguistics domain.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

(Folli & Harley, 2013: 112)

In (50), the predicate is non-agentive, since the subject is the internal argument, and it is affected. This contrasts with the same NVE in combination with *dare* (51), which stand in the same relation as *give* and the unaccusative *get* in English (Harley, 2002).

(51)   a. Gianni ha preso una sgridata

Gianni has.AUX taken.PART a reprimand.

'Gianni has been scolded'

b. Ha dato una sgridata a Gianni

Has.AUX given.PART a reprimand to Gianni

'He/she has scolded Gianni'

In English, not all LVCs with *take* are unaccusative. In fact, the LVCs under analysis in this dissertation have an agentive nature, as the instances in (52)[47]. Therefore, take-LVCs cannot be analysed as the unaccusative variant proposed by Folli & Harley (2013) for some instances with prendre in Italian (in 51).

(52)   a. take {step, picture, walk}

b. {Abstand, Rücksicht, Einsicht} nehmen

{distance, consideration, insight} take

c. prendre {nota, represàlia, decisió}

Take {note, reprisal, decision}

d. tomar {determinación, descanso, fotografía}

Take {determination, rest, picture}

Sanromán Vilas (2017) provides a syntactic and semantic analysis of the verb *tomar* ('take') in both its full (53a) and light use (53b). She concludes that the LV shares

---

[47] A more detailed description of the LVC lemmas included in the analysis has been provided in Chapter 3.

three lexical features with its full use: (i) the centrality of the subject, (ii) the spatial deixis and (iii) inceptive meaning.

(53)   a. *tomar un libro* ('to take a book'), *tomar pastillas* ('to take pills')

b. *tomar la siesta* ('to take a nap'), *tomar una foto* ('to take a photo'), *tomar una decisión* ('to take a decision'), *tomar venganza* ('to take revenge')

More specifically, in the heavy/full use of the verb the subject takes control of the action, there is a movement towards the subject, and it describes a telic event, as in (54).

(54)   Tomó un libro de la estantería

Took.3SG a book from the shelf

'He/she took a book from the shelf'

In contrast, when the verb is part of a LVC, if there is no specific control of the subject, there is always an action that takes place within the subject (Sanromán Vilas, 2017), that is, that they affect the subject (55a). The spatial deixis is related to the movement or location of the object, which moves towards the subject also in a metaphorical way (55b). Finally, the aspect is related to the initial phase of the action or process (55c). Colominas (2001: 210) also describes the LV *prendre* in Catalan as an agentive and telic verb.

(55)   a. La sustancia empezó a tomar forma

The substance started.3SG to take form

'The substance started to take shape'

b. Se tomó una ducha

SE took.3SG a shower

'He/she had a shower'

c. La niña le tomó afecto   (Sanromán Vilas, 2017: 263)

The girl him.DAT took.3SG affection

'The girl became fond of him'

Similarly, Norvig & Lakoff (1987) study the different meanings of *take* in English. As the default semantic representation of the verb, they define *take* as "A receives P". Prototypically, *take* presents a relation between the agent (A) and the patient (P) where the agent receives the patient as a result of taking, as in (56a), which can also be a more metaphorical receiving, that is, perceiving in its light use (56b).

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

(56)   a. The baby took the boy from his mother   (Norvig & Lakoff, 1987: 198)

b. take a look, take a glance

Another relevant property, which has already been observed in the Spanish *tomar,* is related to the movement and spatial deixis, which can be physical in its full use (57a) or more metaphorical in its light use (57b). Finally, the verb *to take* is aspectually related to perfective short actions (58a), while it does not combine with imperfective nominals (58b).

(57)   a. John took the book to Chicago   (Norvig & Lakoff, 1987: 200)

b. I took a punch at John

(58)   a. take a look, take a glance, take a taste, take a whiff

b. *take a stare, *take a leer

In sum, properties shared among all LVs *take* in the Romance and Germanic languages are the following: a metaphorical spatial deixis, the action situated within the subject and certain constraints on aspect. The latter presents some differences between language families: in Spanish and Catalan, *tomar* and *prendre* express the initial phase of an action, while in English and German, *take* and *nehmen* are related to bounded events.

In brief, apart from the aspectual idiosyncrasies of the verb, the main property distinguishing the LV *take* from the other two LVs under study is the fact that the subject in these constructions is both the agent and beneficiary/receiver of the event.

5.4. Interim summary

This chapter has presented an overview of previous approaches to LVCs. First, the status of LVs within the whole construction has been reviewed. The consensus of previous studies has shown that there is no need for two lexical entries for the verb as a full verb (in non-light contexts) and as a LV, since the main difference lies in the context in which the verb is inserted.

From a neoconstructionist approach to argument structure, the LV is analysed as inserted in little *v*. Due to its insertion in this functional layer, the verb can contribute to the event structure of the LVC as a whole, but it does not contribute conceptual semantics because of the lack of a root. The analysis of LVCs is intrinsically related to the analysis of unergative verbs as underlying transitive (ala Hale & Keyser, 1993, 2002), where the former select a DP and the latter involve a root.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 5. Previous approaches to LVCs

When the three LV selected in the present dissertation are taken into consideration (*give*, *take* and *make,* and their corresponding verbs in the languages under study), they are all eventive dynamic verbs with some differences regarding their aspect and argument selection.

With respect to their aspectual properties, *make* is generally related to accomplishments, while *give* and *take* to achievements. The aspect of the whole construction is, however, dependent on the context and the verb does not completely determine its aspect.

As for the argument selection, *give* and *make* can select either one or two arguments, in a ditransitive construction where the IO is introduced by an applicative. The main difference detected between these two LVs is related to their semantics: make is the prototypical creation verb, while give tends to highlight the transfer of the argument. In contrast, take can only select one argument, which can be both the agent and the receiver of the event.

Finally, some differences have been signalled regarding the selection of NVEs introduced by a determiner and bare NVEs, especially in Spanish. The different properties of each kind of nominal have been previously explained through different levels of projection within the nominal domain, namely nP for bare nouns and DP for determined NVEs (as in Kornfeld, 2004). This will be further developed in the typology proposed in Chapter 6.

Chapter 6. A typology of LVCs in Germanic and Romance languages

After the description and analysis of the empirical data in previous chapters, in this section I will present a typology of LVCs in the two Germanic and the two Romance languages under investigation with focus on the NVE and its tight link with the LV.

First, a brief overview on the theoretical approaches to the nominal domain will be provided in §6.1, as well as a more specific coverage of the main difference between nominals in Germanic and Romance languages (§6.1.1).

Second, I will present the argument properties of the different kinds of nominals in LVCs in §6.2, with focus on the properties of NVE introduced by a determiner and bare nouns.

Third, I will put forward an analysis within a neo-constructionist approach of the different properties of nominals after revising previous approaches to the nominal domain in §6.3.

Fourth, I will present a typology with the main difference between nominals introduced by a determiner and bare nominals in §6.4. Three groups are identified: type 1 includes nouns introduced by a determiner with optional modification (§6.4.1), type 2 contains bare nominals (§6.4.2) and type 3 corresponds to an exceptional case of noun incorporation in Rioplatense Spanish (§6.4.3).

Finally, I will summarize the proposed typology in §6.5.

6.1. Theoretical approaches to the nominal domain: a brief overview

Since the difference between nominals introduced by a determiner and bare nominals is key for the typology proposal for LVCs, I will briefly review the most relevant theoretical approaches to the nominal domain. The DP- vs. NP-Hypothesis ongoing debate will be discussed with special focus on the most recent revisions by Bruening (2020) in favour of NP-Hypothesis, and the crosslinguistic revision of Barrie et al. (2021) and Salzmann (2020), which conclude that there is a need of a functional layer above N (similarly to the DP-Hypothesis) to explain some phenomena more accurately (i.e. idiom formation or hybrid nouns in Serbian-Croatian-Bosnian).

The basic question revolving around theoretical approaches to the nominal domain is whether the DP dominates the NP, as in (1), or it is the NP that dominates the determiner system, as in (2). In the first case, the DP-Hypothesis is based on the functional category DP dominating the lexical layer NP, while the NP-Hypothesis proposes that the head is N. Despite the success of the first proposal since Abney (1987), the DP vs. NP debate is still ongoing.

(1)
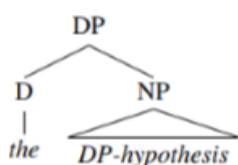
(Blümel & Holler, 2022: 2)

(2)
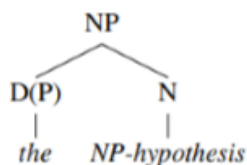
(Blümel & Holler, 2022: 2)

The first proposal of a DP analysis of the noun phrase was put forward by Abney (1987), who bases his proposal on the advances in X'-theory in Chomsky's *Barriers* (1986). From then, the structure of the VP allows functional elements which project other than the verb, that is, Complementizer Phrase (CP) and the Inflection Phrase (IP). These constitute the extended projection of the verb (Grimshaw, 1991; cf. deMena Travis, 2014). When this is applied to the nominal domain, the determiner phrase (DP) is the maximal projection by determiner elements which heads the noun phrase.

The syntactic evidence put forward by Bernstein (2001: 358) for the DP-hypothesis is that the internal structure of the phrase is parallel to the proposal for the verbal phrase: the external argument of the DP is generated in SpecNP and the internal argument as a complement of N, as proposed by Ritter (1988) on Hebrew noun phrases. However, there is some debate as to whether these arguments raise from N-to-D, and to which specifier position they do (as in Picallo, 1994, for possessive pronouns in Catalan).

The semantic evidence for the DP analyses is that the presence of a determiner provides a noun phrase with argument properties, as discussed in Longobardi (1994: 619)

for Italian bare nouns, as in (3). The bare nominal in (3a) is a predicate of the copula and corresponds to an NP-predicate, which cannot serve as the head of a relative clause; whereas in (3b) the same nominal is introduced by an indefinite article and is, thus, an argument which can serve as a head of the relative clause introduced by *che* ('that').

> (3) a. Gianni é medico (*che...)
>
> b. Giani é un medico (che...)

Due to the existence of languages without articles, the universality of the DP projection has been under question. Since its definition, the DP Hypothesis has been widely debated (Blümel & Holler, 2022), some approaches deny the need of the D functional head for nominal expressions (as in Bruening, 2009, 2020); whereas others state that previous arguments (e.g. Bernstein 2001, Longobardi 2001, Salzmann 2020) remain inconclusive.

In fact, Bruening (2009, 2020, 2022) argues against the DP-Hypothesis by showing discrepancies between nominals and clauses to support the NP-Hypothesis. Bruening's analysis does not lack functional heads within the noun phrase, but he proposed that these are dependents of the N projection, as represented in (4). In this analysis, the demonstrative (Dem), determiner (Det) and numeral (Num) are comparable to adjuncts like the adjectival phrase (AP), because they are dependent of the head N.

> (4)
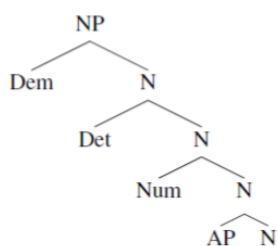
(Bruening, 2020: 2)

With evidence from conventionalized fixed expressions (or idioms), Bruening et al. (2018: 17) show that in the majority of verb-object expressions, the V and N are fixed, while the D can vary, as in (5).

> (5) a. bark up the wrong tree – You're barking up another wrong tree.
>
> b. beat around the bush – Let's beat around this bush no more.
>
> c. rock the boat – This'll rock some boats.

In light of these examples, the verb selects an NP headed by N, and the other branches within this NP (determiners, numerals and classifiers) can be open, as they can vary even in fixed expressions. For Bruening et al. (2018), determiners in such idioms are optional modifiers, and they do not behave as other functional elements such as the preposition in PPs. This is taken as proof that the head of the phrase must be the N, as the verb does not select a specific determiner, as exemplified in (5).

However, Bruening et al.'s analysis of fixed expressions in which functional material (i.e. determiners) does not participate in idiom formation poses several problems, which are pointed out in Stone (2016) and Barrie et al. (2021).

Among others, if the determiner is not involved in building the idiomatic meaning, the interpretation of idioms chunks replaced by a pronoun cannot be accounted for. In (6), the reading of the idiom *spill the beans*, that is, *revealing a secret* is still maintained when the whole noun phrase (*the beans*) is replaced by the pronoun (*them*). Therefore, the determiner must be part of the idiom.

(6) John spilled the beans, but Mary didn't spill *them*. (Barrie et al. 2021: 224)

In their work, Barrie et al. (2021) argue in defence of the DP analysis of the nominal domain with crosslinguistic evidence. They assume that there is another layer above DP (namely KP), as in (7), which shows that there is no local relationship between V and D, as it is impossible for a verb to select the definiteness of the nominal (as argued previously by Bruening et al., 2018).

(7)
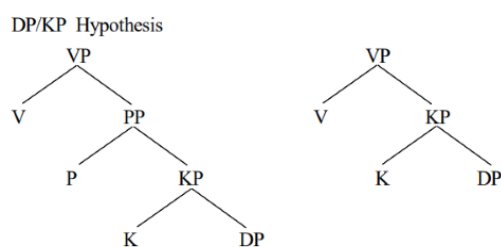
(Barrie et al., 2021: 212)

This analysis also supports the selection of bare nominals: even though it assumes that verbs can select full KPs, they can also select below this maximal functional layer (i.e. NP or nP). In Barrie & Matthieu (2016), it is proven that there are nominal complements of different sizes which can be selected by the verb (i.e., Oneida selects an

nP, and Niuean a NumP). Therefore, the need for a DP hypothesis is supported by the variety of nominals which can be selected by the verb, which need the functional layers above NP to be legitimated.

In turn, Salzmann (2020) reviews the arguments that have been provided in favour of both approaches (the DP-hypothesis and the NP-Hypothesis) and finds them inconclusive for both theories. An argument that Salzmann considers to be key in this debate is related to the notion of headedness: "The features of the head are present on the maximal projection. This implies that the head is closer to outside probes than other constituents of the noun phrase. This is expected to be visible in selection and agreement processes." (Salzmann, 2020: 41).

In the case of agreement, the DP and NP hypothesis make different predictions with data from hybrid nouns in Bosnian-Croatian-Serbian data. The agreement switches between grammatical and biological gender agreement in these nouns show that the features of D must be closer to V than the features of N. Hence, such agreement can only take place if D is a head above the NP, as in the DP-Hypothesis. This is exemplified in (8), where the noun *vladika* ('bishop') is grammatically feminine but semantically masculine and, therefore, there is optionality in the selection of gender, which gives rise to a gender mismatch between the determiner and the adjective.

(8) On-i star-e vladike su se posvadjal-i/*posvadjal-e na ulici.

(Salzmann, 2020: 34)

those-**m**.pl old-**f**.pl bishops are refl argued-**m**.pl/argued-**f.**pl on street

'Those old bishops argued on the street.'

In sum, although both the DP- and the NP-Hypothesis are considered to be able to account for several cross-linguistic nominal phenomena, there is evidence from different languages (as pointed out in Barrie et al., 2021; Salzmann, 2020; among others) that the functional layers above the nominal head N can account for more issues related to the divergences between the determined NVEs and bare nominals. In the proposal presented in this dissertation, I will assume that the DP-Hypothesis and the extended projection within the nominal can explain some of the restrictions found in the NVE in the corpus database (described in more detail in Chapter 3).

6.1.1 Nominals in Germanic vs. Romance languages

The nominal domain presents certain differences between Germanic and Romance languages, especially in the object position. Their main divergence lies on the possibility of having bare nouns in that position. The possibility of their cooccurrence with nouns introduced by determiners (or full DPs) has posed a theoretical challenge for linguistics.

In general, Germanic languages accept bare plurals in object position while bare singulars are restricted to mass nouns (9). In these languages, bare plurals have kind reference (that is, they denote the sum of all realizations of the kind), and mass nouns are also inherently plural as they have kind reference. In fact, English bare plurals are assumed to be DPs (Carlson, 2003, and references therein) while bare singulars are analysed as NPs.

(9) a. They wanted dogs. *vs* *They wanted dog.

    b. They drank wine.

In contrast, Romance languages can have bare count singulars in object position, as in (10), although these contexts are restricted to certain verbs, the so-called *have*-predicates (Borthen 2003; Espinal & McNally, 2011).

(10)   a. Juan tiene coraje.   (Spanish)

        Juan has courage.

        'Juan is brave'

        b. Sempre porta barret.   (Catalan)

        Always wears hat

        'he/she always wears a hat'

This divergence is already discussed in Chierchia (1998)'s proposal of the nominal parameter, as illustrated in Table 59. In his ontology, DPs are considered to be arguments crosslinguistically, while NPs are different kind of predicates in every language. The prediction of Chierchia's approach is that languages with argumental nominals (like Chinese) will freely allow them in argument position but not as predicates, while languages with predicative nominals (like French) will need a Determiner in order to be interpreted as arguments. Finally, English shows a mixed behaviour allowing nominals in different positions when certain semantic operations apply.

Language type	[+arg, -pred]	[-arg, +pred]	[+arg, +pred]
	Chinese	French	English
Noun Type	<e>	<e,t>	<e> or <e,t>

Table 59. Types according to the Nominal Mapping Parameter (in Espinal, 2013: 68)

However, this three-way partition does not accommodate other Romance languages but just Modern French (as pointed put in Dobrovie-Sorin et al., 2006; Espinal & McNally, 2011; Espinal, 2013).

In the case of Romance languages, the different types of nominal expressions in object positions are summarized in Table 60. The main property that distinguishes object bare count nominals is that they denote properties of kinds (Bosque, 1999) and are unmarked for number and definiteness (Espinal, 2013; Espinal & McNally, 2011).

Reference	Representation	Syntactic correlate
Entity	<e>	DP
Properties of kinds	$<e^k, t>$	NP in object position of *have*-predicates
Properties of objects	$<e^0, t>$	NumP (NPs specified for number): Romance bare plurals and indefinites

Table 60. Types for nominal expressions in object position (Espinal, 2013: 74)

Since only DPs can be arguments (Longobardi, 2001), there have been several proposals to reconcile bare nominals with argumenthood, especially pointing to the assumption that there is a null D (Borer, 2005). Longobardi (1994) interprets this null D as a default existential Quantifier. In fact, bare objects show reduced referential features: discourse anaphoric pronouns do not find an antecedent in the bare noun, as in (11), where the bare singular is not argumental because it must refer to the event of a music performance (Le Bruyn et al., 2017).

(11)   a. Phil is playing piano$_i$ for the choir. #He complains it$_i$ is out of tune.

         b. *Tomé nota y la dejé en la mesa.   (Montero Gálvez, 2014)

         'I took note and left *it on the table'

After this brief overview of the theoretical approaches to the nominal domain, the characteristics of the nominals in LVCs will be discussed with a special focus on the role of bare nominals.

6.2. LVCs at two ends of a continuum

As the present study seeks to find regularities and dissimilarities in the syntactic-semantic structure of LVCs with a focus on the nominal element, I will review the previous analyses of LVCs with a focus on the NVE.

Interestingly, as in Acedo-Matellán & Pineda (2019)[48], our data show that there is a continuum in the degree of cohesion between the light verb and the NVE, from transitive-like structures (12) to more fixed LVCs with bare nouns (13).

(12)   a. The committee *made a decision* during the meeting.

b. Bemis Anwälte hätten ihm *den Rat gegeben.*   (DWDS, 2000)

'The Bemis' lawyers would have given him the advice'

c. Garralda somriu i li *dóna una empenta* per fer-lo sortir del despatx.

(CTILC, 2004)

'Garralda smiles and gives him a push to lead him out of the office'

d. Levantó la cámara y *tomó una fotografía.* (CORPES XXI, 2001)

'He/she lifted the camera and took a picture'

(13)   a. No man knew to whom he ought to *make payment* first. (COCA, 2012)

b. Es darf uns nicht egal sein, wie die EU auf Standards *Einfluss nimmt*,

(DWDS, 2019)

'It should not matter to us, how the EU influences the Standards'

c. Era un whig o lliberal moderat que *donava suport* als drets civils de les classes més baixes   (CTILC, 2009)

'He was a whig or a moderate liberal who supported the civil rights of lower classes'

d. En el último tiempo *hemos tomado conciencia* de eso.

(CORPES XXI, 2001)

'We have lately become aware of it'

---

[48] In Basque LVCs, there are transitive-like structures which select the auxiliay *edun* with an ergative subject and the NVE is the direct object, which can be separated from the verb and bear partitive case (i.e. *dantza egin* 'do dance' or *negar egin* 'do weep'). At the other end of the continuum, there are fossilized LVCs as instances of incorporation, which do not admit partitive case and present adjacency of the NVE to the verb, as well as have a single phonological word (i.e. *hitz egin* 'do word' meaning 'to speak' or *alde egin* 'do side/zone' meaning 'to leave'). There is also a group of Basque LVCs with an ambiguous behaviour between both extremes (Martinez, 2015, p. 337-373).

Chapter 6. A typology of LVCs in Germanic and Romance languages

Some differences between Germanic and Romance languages have been found in this regard. The majority of English LVCs are transitive-like with determiner phrases as the NVE (14) whereas most LVCs in Spanish can accept the NVE with and without a determiner (15).

(14)   take *(a) break, give *(a) description, make *(an) impression

(15)   tomar (una) nota (lit. 'take (a) note'), hacer (una) pausa (lit. 'make (a) break'), dar (una) ayuda (lit. 'give (a) help')

From the LVC database conforming the empirical basis of this dissertation, there is a clear distinction between English and Spanish LVCs: while 75.61% of Spanish LVCs and 70.73% of Catalan LVCs accept the construction with an optional determiner, in English 71.43% have an obligatory determiner. This restriction could be explained through the different determination possibilities in both languages since Romance languages have more flexibility with bare nouns[49] in object position than English (Longobardi, 2001). However, this tendency is not applicable to the Germanic language family, as German LVCs show the opposite tendency to English because it has 79.31% of LVCs with an optional determiner (see Table 61 and Appendix 2).

	English	German	Catalan	Spanish
Determiner is optional *make (a) conversation*	12/42 (28.57%)	23/29 (79.31%)	29/41 (70.73%)	31/41 (75.61%)
Determiner is obligatory *give *(a) hug*	30/42 (71.43%)	6/29 (20.26%)	1/41 (29.27%)	10/41 (24.39%)

Table 61. Distribution of obligation and optionality of a determiner in the LVCs under analysis

The presence (or lack) of determination is crucial for the syntactic structure in LVCs cross-linguistically. Some English LVCs behave like regular transitive verbs, as they can be passivized (16b) and pronominalized (16c); whereas others do not allow these tests: passivization (17b) and pronominalization (17c) are not acceptable, where the reference is a bare nominal within the LVC (*conversation*).

(16)   a. The committee made a decision during the meeting.

        b. The decision was made during the meeting.

        c. She made a decision and made it loud and clear.

---

[49] In this chapter, the term bare noun/s will refer to bare singulars, that is, nouns in singular not introduced by any kind of determiner. Bare plurals are not included in this group because they are considered to have a D layer introduced in their syntax due to the plural marker (Cyrino & Espinal, 2019).

(17) a. If you've ever made conversation with someone on a plane.

b. *(The) conversation was made on a plane.

c. ?He made conversation and made it very enjoyable.

Determination is also relevant in the syntactic structure of Spanish LVCs: the NVE introduced by a determiner can be passivized (18b) and pronominalized (18c) as a transitive-like construction. There are, however, certain divergences with the English data in relation to the LVCs with a bare noun: passivization is not allowed in periphrastic passives (19b) but it is possible with *se*-passives where the NVE is postverbal (19c), which is a property of *se*-passives in Spanish (as noted in López, 2002; Fábregas, 2021).

(18) a. El Ministerio tomó una decisión sobre el asunto.

'The Ministry took a decision about the matter.'

b. La decisión fue tomada con cautela,   (CORPES XXI, 2001)

'The decision was taken with caution'

c. Había tomado una decisión y la llevaría a cabo,

(CORPES XXI, 2001)

'She had taken a decision and would carry it out'

(19) a. Tomó nota en su libreta de bolsillo.   (CORPES XXI, 2002)

'He took note in his pocket notebook'

b. *Nota fue tomada en su libreta de bolsillo.

'Note was taken in his pocket notebook'

c. Se tomó nota en su libreta de bolsillo.

SE took.3SG note in his notebook of pocket

'Note was taken in his pocket notebook'

Likewise, pronominalization of the bare NVE is generally not acceptable, but our sample shows instances where a bare noun can be pronominalized when the reference is present in the immediately previous clause (20). Such instances contrast with Mendívil-Giró's (1999) judgments where the only possible pronoun for LVCs with bare nouns in Spanish is the neuter pronoun *lo,* in (21), which refers to the whole predicate *tener manía* ('have hostility') rather than the NVE itself.

(20) Si necesitan ayuda, se la daremos,   (CORPES XXI, 2008)

If   need.3PL help, them.DAT it.ACC give.FUT.1PL

'If they need help, we will give it to them.'

(21)   Tengo manía a Luis y no {*la/lo} puedo olvidar

(Mendívil-Giró, 1999: 277)

'I have hostility towards Luis and I cannot forget it'

A similar tendency is also found in Catalan LVCs which allow passivization in determined NVE (22a)[50] as well as pronominalization (22b). In the case of bare NVEs, they cannot be passivized (23b) unless it is with the pronominal passive equivalent to *se*-passive (23c), which is the most common form of passivization in Catalan (as observed in Wheeler et al., 1999: 505).

(22)   a. La bufetada havia estat donada i assumida.   (CTILC, 2003)

'The slap was given and accepted'

b. El comte va donar el permís a l'abat, però el va donar a contracor.

'The Count gave permission to the abbot, but he gave it reluctantly'

(23)   a. un dia pren consciència d'aquesta realitat transcendent. (CTILC, 2009)

'one day, he/she becomes aware of this significant reality'

b. *Consciència és presa d'aquesta realitat transcendent. / *Consciència d'aquesta realitat transcendent és presa.

'Awareness is risen of this significant reality. / Awareness of this significant reality is risen'

c. Es pren consciència d'aquesta realitat transcendent.

They also present restrictions with regard to pronominalization: they accept the partitive *en*, but they tend to reject definite pronouns (24). Espinal & Dobrovie-Sorin (2006: 15) show that Catalan bare nominals can be pronominalized by both kinds of pronouns but there is a semantic distinction: the partitive *en* expresses a type anaphora (i.e. the antecedent is a type of an entity), while that of the definitive pronouns *el/la/els/les* is a token anaphora[51] (i.e. the antecedent is an individual).

---

[50] This contrasts with the proposal in the Catalan standard norm (GIEC, 2016: 23.2.1.1b) which rejects the periphrastic passivization with *ser* ('be') with light verbs.

    (i)   *L'esquena fou donada a la Maria per la Sandra.
       'The back was given to Maria by Sandra'

[51] They focus on instances that are not LVCs, but rather bare nominals in *have*-predicates as in (i).
    (i) Vol portar corbata per portar-ne
    wants to.wear tie to wear-CL.gen
    'He wants to wear (a) tie to wear it'
    b. Vol portar corbata per lluir-la
    wants to.wear tie to show.off-CL.ac.fem.sg
    'He want to wear (a) tie to show it off'

(24)   És un espai per donar acollida als sense sostre, però només en/#la dona de 7:45 a 18:00.

'It is a place to give shelter to homeless, but it gives some/#it from 7.45 to 18.00'

In German LVCs, determined NVEs can be both passivized (25b) and pronominalized (25c). As seen in Chapter 3, German LVCs showed a tendency to co-occur with bare nominals, which present similar restrictions as those of the other languages under study: it rejects passivization with stative *sein*-passives (26b), but it accepts eventive *werden*-passives (26c)[52]. They cannot be pronominalized, either (27).

(25)   a. Van der Vaart aber hat bereits die Antwort gegeben: (DWDS, 2005)

'But Van der Vaar has already given the answer'

b. Die Antwort ist gegeben.

'The answer is given'

c. Hat Van der Vaart die Antwort gegeben? Er hat sie endlich gegeben.

'Has Van der Vaart given the answer? He has finally given it/one'

(26)   a. Vielleicht würden sie bald Abschied nehmen.   (DWDS, 2002)

'They might say goodbye soon'

b. *Abschied ist bald genommen.

'Goodbye is soon said'

c. Am Mittwoch wurde im Roten Rathaus Abschied genommen.

(DWDS, 2005)

'Farewell will be taken on Wednesday in the Roten Rathaus'

(27)   a. Alle Gäste hatten Platz genommen.

'All guests have taken a seat'

b. *Er hat auch einen genommen.

'He has also taken one'

So far, only two tests have been discussed regarding the argument properties of the NVE: passivization and pronominalization. Both have shown that determined

---

[52] As native speakers have noted, it can appear in impersonal *werden*-passives (i), where the subject is an expletive *es* and the bare nominal remains in the object position; which contrasts with the instances found in the corpus where the subject is omitted, as in (26c).
(i)   Es wurde Platz genommen.
      'A seat was taken'

nominals and bare nominals have divergent behaviours. In this line, other tests such as *wh*-question formation, relativization and topicalization present similar results establishing a clear difference between determined and bare NVEs.

With regard to *wh*-question formation (Kearns, 1988; MacFarland, 1995: 61), only determined NVE can be extracted to answer an interrogation of the predicate (28) while bare nouns do not have the entity to stand on their own cross-linguistically (29)[53].

(28)   a. What did he give? A groan.

b. ¿Qué le hicieron? Un masaje.

'What did they give him? A massage.'

c. Què va donar-li? Una pallissa.

'What did they give him? A beating'

d. Was hat er dir gegeben? Die Antwort.

'What did he give you? The answer'

(29)   a. What did they make? *Conversation.

b. ¿Qué dieron? ?/*Permiso a los niños.

'What did they give? Permission to the kids.'

c. Què van prendre? ?/*Consciència de la importància del canvi climàtic.

'What did they take? Awareness of the climate change relevance'

d. Was haben die Gäste genommen? *Platz.

'What have the guests taken? Seat.'

Regarding relativization, MacFarland (1995) observed that the nominal element of LVCs can select a non-restrictive relative clause (contra Kearns, 1988). The examples provided are, however, restricted to determined NVE, as in (30).

(30)   a. The man on the right gave a loud groan, which chilled me to the marrow

(MacFarland, 1995: 75)

b. Márgara Locquolo le daba respuestas que a su vez eran preguntas.

(CORPES XXI, 2001)

---

[53] Native speakers accept bare nominals in this position in echo (or pseudo echo) wh-questions, as in (i).
(i)   A: Die Gäste haben Platz genommen.
'The guests have taken (a) seat'
B: Sorry, was haben die Gäste genommen?
'Sorry, what have the guests taken'
A: Platz.
'(A) seat'

'Márgara Locquolo gave him answers that were questions at the same time'

c. com si en fes una fotografia que podrà tornar-se a mirar tota la vida.

(CTILC, 2014)

'as if s/he could take a picture which s/he could look again her whole life'

d. weil jeder der Beteiligten Fehler macht, die er später bereut.

(DWDS, 2000)

'because each participant makes mistakes which he later regrets'

In (31), the instances of LVCs with bare nominals with a relative clause are acceptable, but only if the scope of the clause is the whole LVC. When the relative clauses are built to refer only to the NVE, their acceptability is only possible when a determiner is included (32).

(31)   a. I made conversation with a stranger, which is very unlike me.

b. Los magnates dieron ayuda, que al principio parecía bienintencionada.

'The tycoons gave help which at first seemed to be well-meaning'

c. La institució va donar consell als socis, que després ningú no va valorar.

'The institution gave advice to the members which later nobody appreciated'

d. Auch in dieser Talkshow hat das AfD-Gespenst Platz genommen, das seit Monaten alle politischen Debatten duschwebt.   (Google)

'Also in this Talkshow, the AfD-apparition has taken a seat, which has interweaven all political debates for months'

(32)   a. I made *(a) conversation with a stranger, which was a very interesting one.

b. Dieron *(una) ayuda a los refugiados que se criticó mucho en los medios.

'They gave help to the refugees which was much criticised in the media'

c. J. M. va donar *(un) consell que es va recordar a totes les reunions de l'empresa.

'J. M. gave advice which was recalled in all company meetings'

d. Das Publikum hat an den runden Tischen *(einen) Platz genommen, das ein oder andere Kaltgetränk bereitstand.

'The audience has taken a seat in the round table, which have one or other
cold drinks ready'

Finally, focus fronting also shows differences between the determined and bare
nominals in Romance languages, since it is allowed only when some determiner
introduces the object (33), but it is not accepted with bare nominals (34).

(33)   a. La misma decisión tomó Pablo.

'The same decision took Pablo'

b. Un consell va donar el mossén, no una opinió.

'A (piece of) advice gave the priest, not an opinion'

(34)   a. *Conciencia tomaron todos juntos.

'Awareness they took altogether'

b. *Permís van donar als estudiants.

'Permission they gave to students'

A parallel behaviour is found in the case of dislocation with clitic doubling in the
two Romance languages: determined NVEs allow it (35), while bare nominals disallow
it both in Spanish and Catalan (36a,b), although the Catalan bare nominal can be
dislocated when the partitive pronoun is used in the doubling (36c).

(35)   a. Esta crítica la hizo el jesuita Danièlou,   (CORPES XXI, 2002)

'This criticism was done by the Jesuit Danièlou'

b. La resposta la donarà ell mateix a Carles V, uns anys després:

(CTILC, 2009)

'The answer will be given by Carles V himself, some years later'

(36)   a. *Crítica la hizo el jesuita Danièlou,

'Criticism was done by the Jesuit Danièlou'

b. *Resposta la donarà ell mateix a Carles V

'Answer will be given by Carles V himself'

c. (De) Resposta ja en van donar ell mateix

'Answer was already given by himself'

In contrast, data from English and German differ because English disallows it in
any case (37), whereas German allows it in both kinds of NVEs (38) (cf. Müller, 2022;
among others).

(37)   a. *Some notes, they took.

b. [?]A wonderful answer, Maria gave.[54]

c. *Advice, the President gave.

(38)   a. Die Erlaubnis geben nicht mehr behandelnde Ärzte. (DWDS, 2003)

'The permission is given by no longer treating doctors'

b. Anstoß nehmen Gesundheitsexperten nach einem Bericht

(DWDS, 2013)

'Offense take experts after a report'

For Slabakova et al. (2012), it is difficult to compare the Romance constructions with the English topicalization: the strategy for the latter, when new information is not the final element of the clause, is stress shift (Reinhart, 2006). Therefore, instances in (37) are only possible when the prosodic stress is placed in the fronted element.

In contrast, German seems to be more flexible with topicalization and fronting in general. In V2 languages, all constituents can be fronted with some exceptions, such as reflexive pronouns, expletive objects and certain particles (Müller, 2022: 19). In fact, the fronted element can be of any grammatical function, and it can cross several clause boundaries (see Müller, 2022: 22-23; Höhle, 1982; Krifka, 1998).

In sum, most tests on the argument properties of the NVE show a clear distinction between the determined and bare nominals which hold across languages (Table 62). Only NVE introduced by a determiner can be passivized, pronominalized, interrogated and accept relative clauses with scope on the nominal element in all four languages, whereas bare NVEs pose more restrictions: they cannot be passivized, although they are accepted in *se*-passives in Catalan and Spanish; they cannot be pronominalized, except through partitive *en* in Catalan; they cannot be interrogated, unless it is in an echo-question; and they do not accept relatives clauses which have scope on the nominal element itself, although they can select relatives with scope on the whole LVC. In contrast, the behaviour regarding topicalization differs across languages: English does not accept it in any case, while German allows it for both determined and bare nominals. They contrast with the Romance languages for which it is acceptable with determined NVE and usually not acceptable when the topic is a bare noun, except through partitive *en* in Catalan.

---

[54] A native speaker suggests that this particular example could be accepted as part of oral speech.
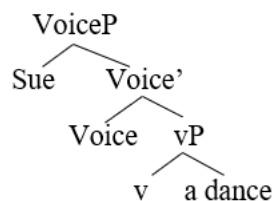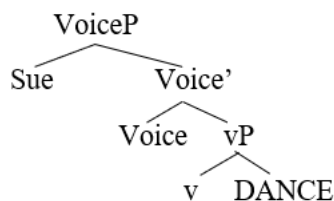
TESTS		English	German	Catalan	Spanish
Passivization	Det	✓	✓	✓	✓
	Bare	*	Only with *werden* passives	Only with *se*-passives	Only with *se*-passives
Pronominalization	Det	✓	✓	✓	✓
	Bare	*	*	✓ (with partitive *en*)	*
Question Formation	Det	✓	✓	✓	✓
	Bare	*	*	*	*
Relatives	Det	✓	✓	✓	✓
	Bare	*	*	*	*
Topicalization	Det	*	✓	✓	✓
	Bare	*	✓	✓ (with partitive *en*)	*

Table 62. Tests on argument properties of the NVE in LVCs

Like other non-light constructions, LVCs with a determiner in the nominal display properties of transitive-like constructions, and they should be analysed as such. In contrast, when the NVE is a bare noun, it does not behave as a canonical argument, and can thus not be analysed as such. There is a need to find an analysis which can account for both types of LVCs.

6.3. Previous analyses within a neo-constructionist approach

The proposal of analysis in this dissertation is based on the neo-constructionist approach to argument structure developed in Acedo-Matellán, (2010, 2016), which builds on Hale & Keyser (1993, 2002 and much related work), where unergative verbs are analysed as an underlying transitive structure with a light verb (repeated here as 39). The difference between unergative verbs like *dance* and its corresponding LVC (*do a dance*) lies in the nature of the complement within the structure: in (39a) the complement is the root DANCE, whereas in (39b) the complement is a DP in a transitive construction.

(39)   a. Unergative *danced*   b. Transitive creation *did a dance*

VoiceP
  Sue   Voice'
         Voice   vP
              v   DANCE

VoiceP
  Sue   Voice'
         Voice   vP
   v   a dance

(Acedo-Matellán, 2016: 31)

In the same line, Myler's proposal (2016: 24) is that a *have*-LVC contains a v but no root, which can be explained by the LV contributing event structure (being the realization of v) but no conceptual semantics (encapsulated in the root) to the LVC.

In (39b) the analysis is that of a transitive LVC where the verb *do* is followed by a DP *a dance* which behaves like a canonical argument (as seen in the tests for argumenthood in Table 62). If we argue that LVCs should be analysed as other verbo-nominal constructions, the analysis here provided is valid for transitive-like LVCs, which lie at one end of the continuum.

At the other end of the continuum, the degree of cohesion between the LV and the NVE is higher, and the nominal element does not behave as a canonical argument of a transitive construction (as seen in §6.1). In these cases, the NVE is not introduced by a determiner, so it cannot be a DP. Our proposal is that the main structure remains the same (43b), but the complement is a nominal phrase with different levels of projection depending on the modification selection, which cannot behave as a canonical argument because it does not project the functional layer D (Longobardi, 2005).

This proposal is in line with Kornfeld's proposal (2004) for the analysis of Spanish LVCs. The distinction between LVCs with bare nominals (as in *dar clase*) and expressions with a DP (as in *dar una clase*) lies in the kind of phrases the light verb combines with: an *n*P or a DP, as well as the level of introduction within the tree (v^0 vs. *v'*). For Kornfeld (2004), the lack of D in the first case is enough to trigger the complex predicate analysis, since only DPs can behave as arguments of a phrase and, thus, the bare nominal (*n*P) has to merge at v^0 in order to be legitimized as a constituent, as shown in (40).

(40)

a. vP

b. vP
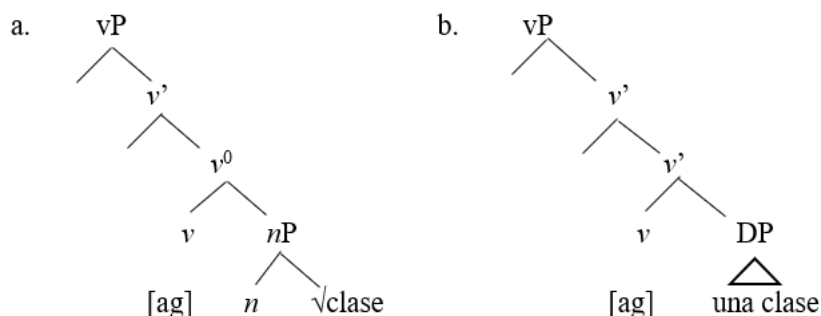
(Adapted from Kornfeld, 2004: 220)

More generally, the idiosyncratic behaviour of bare nouns, which do not behave as regular arguments (as seen in Table 62), has been traditionally attributed to pseudo-incorporation: a family of phenomena that display semantic but not syntactic properties of incorporation (Massam, 2001). In Espinal & McNally's work (2011) on bare nouns in Catalan and Spanish, it is shown that syntactic and semantic selection do not always match perfectly.

When comparing pseudo-incorporated (PI) nominals and LVCs in Persian, Fleischhauer (2020) shows that the main difference between both types of constructions is that the verbal head of an LVC is a light verb, whereas a PI nominal appears with a heavy verb; as the summary of the semantic and morphosyntactic properties of Persian predicational construction types show (Table 63).

Predicational construction type	Noun	Verb
Pseudo-incorporation construction	Semantics: non-referential object argument Morphosyntax: bare direct object noun	Heavy verb determines situation type
Regular (transitive) predicate-argument construction	Semantics: referential object argument Morphosyntax: non-bare direct object noun	Heavy verb determines situation type
Light verb construction	Semantics: *no restrictions on referentiality* Morphosyntax: *no restriction* determines situation type	light verb

Table 63. Summary of the semantic and morphosyntactic properties of Persian predicational construction types (Fleischhauer, 2020: 8).

However, when comparing LVCs with pseudo-incorporation constructions in Persian with *have*-predicates, the nominals in LVCs show no restrictions on referentiality, because not only bare NVEs have been considered in the analysis. For instance, in (41) the LVC *sedâ dadân* ('make (a) sound') the nominal includes an indefinite mark. Therefore, the contrast with bare nouns is not complete.

(41)   Âbgarmkon sedâ-i dâd. Ân (sedâ) boland bud.

        boiler sound-INDEF gave. DEM sound loud be.PST

        'The boiler made a sound. That (sound) was loud.'

(Fleischhauer & Neisani, 2020: 13)

Taking Fleischhauer's approach to Persian LVCs (2020) into account, a comparison between bare nouns in LVCs and in PI constructions will be presented. The tests show that some of the main properties of PI nominals are shared with LVCs, at least in Romance languages (Espinal & McNally, 2011).

In both cases (PIs and NVEs in LVCs), the doubling of the bare noun by a full DP is not permitted. The bare noun occupies the position of the syntactic complement of the verb although it cannot behave as a canonical argument because it is not selected semantically by it (Espinal & McNally, 2011: 89). This test can be applied to both PI nominals in *have*-predicates constructions (42) and LVCs with bare NVE (43).

(42)   a.*Tengo piso un duplex.

        Have.1SG flat a duplex

        b. *Lleva sombrero una pamela.

        Wear.3SG hat   a   sunhat

(43)   a. *Dio respuesta un mensaje.

        Gave.3SG answer a message

        b. *Hicimos análisis un estudio.

        Made.1SG analysis a study

Another property of PI nominals is that they take narrow scope in negation. In (44a) it is understood that the person is not looking for any apartment at all, rather than a specific kind of apartment, as well as in (44b) they are not attending any school at all. In

contrast, (44c) shows that the indefinite article can take either narrow scope or wide scope: it could be that the speaker is not looking for any apartment, or that there is an apartment the speaker is not looking for.

(44)   a. No busco piso   (Espinal & McNally, 2011)

Not look.for.1SG flat

'I am not looking for an(y) apartment.'

b. I don't attend school   (Stvan, 2009)

c. No busco un piso   (Espinal & McNally, 2011)

Not look.for.1SG a flat

'I'm not looking for any apartment. /There is an apartment I am not looking for.'

This contrast can also be found in LVCs allowing bare nouns as the NVE. In (45) the difference is that bare nouns can only take narrow scope in negation, while determined NVEs can also take a wide scope: in (45a), there is no answer at all; while in (45b) it could be that there is no answer at all or a specific kind of answer.

(45)   a. Los presupuestos del 2021 no dan respuesta a la crisis.

The budgets   of.the 2021 not give.3SG answer to the crisis

b. Los presupuestos del 2021 no dan una respuesta a la crisis.

The budgets   of.the 2021 not give.3SG an answer to the crisis

'The budgets do not provide an answer to the crisis. / There is a specific answer that they cannot provide.'

With respect to number, PI nominals in Romance languages are number neutral, according to Espinal & McNally (2011), as in (46). Following Farkas & de Swart (2004), number neutrality is applied to nouns that do not only describe an individual but also they can refer to a group of objects.

(46)   a. Tengo cuenta corriente en el Deutsche Bank.

(Espinal & McNally, 2011)

'I am a client of the DB.' (I may have one account, or more than one)

b. Busco pis.   (Espinal & McNally, 2011)

'I am looking for a flat.' (I may be looking for one flat, or more than one)

Number neutrality is a property that can be applied to the majority of LVCs with bare nouns, as in (47).

(47)   a. El director es quien tiene que dar permiso.

'It is the director who has to give permission.' (It may be one kind of permission, or different kinds)

b. El sustituto toma nota de nuestros consejos.

'The replacement takes note of our advice.' (It may be one note, or more)

The main differences between the properties of PI-nominals (as established in Espinal & McNally, 2011) and bare nominals in LVCs are related to the restrictions in modification and in the semantics of the event.

Regarding modification, PI-nominals only accept kind or type modification (48a); while they are unacceptable with qualitative or descriptive adjectives (48b,c,d). In Spanish, prenominal adjectives are also unacceptable, because they modify individual objects (48e).

(48)   a. Este proyecto posee licencia municipal.   (Espinal & McNally, 2011)

this project possesses permit municipal

'This project has a permit from the city.'

b. Tiene pareja {*alta / *enferma}.   (Espinal & McNally, 2011)

Has   partner   tall   sick

'S/he has sick/tall partner'

c. Attending night school / *crowded school.   (Stvan, 2009)

d. I've left (*busy) town.   (Stvan, 2009)

e. Tiene {*falsa} pareja.

Has   fake   partner

In contrast, bare nouns in LVCs show fewer restrictions with respect to modification, as already pointed out in Fleischhauer (2020) for Persian LVCs, where bare NVEs can select descriptive adjectives. LVCs with bare nominals accept both kind modification (49) and qualitative modification (50), even when the adjective is in a prenominal position (51), although there are certain restrictions (the most acceptable adjective in this position is *bueno* 'good').

(49)   a. El médico da consejo alimentario.

'The doctor gives nutritional advice'

b. Tomé nota escrita de la conferencia.

'I took written note of the conference'

(50)   a. Están preparadas para dar respuesta visible y pacífica a la sentencia.

'They are ready to give a visible and pacific answer to the sentence'

b. Puedes tomar nota breve y telegráfica de la respuesta.

'You can take brief and telegraphic note of the answer'

(51)   a. deberían tomar {buena /*mala} nota de su honestidad.

'They should take good/bad note of their honesty'

b. Cómo hacer {buena / mala} crítica televisiva?

'How to make good/bad television criticism?'

Another property of PI-nominals is that they name a well-established or institutionalized activity (Espinal & McNally, 2011; Le Bruyn et al., 2017) as in (52), but this cannot be applied to LVCs with bare nouns since their combinatory possibilities are higher (53)[55].

(52)   a. Buscar piso, comprar piso, necesitar piso, tener piso, alquilar piso, llevar falda, llevar sombrero, tener pareja, tener coche, tener sombrero, tener falda, comprar coche, alquilar coche, dejar coche, coger coche

b. attend school, leave town   (Stvan 2009)

(53)   a. dar autorización, dar ayuda, dar vuelta, dar paso, dar respuesta, dar clase, dar cambio, dar consejo, dar instrucción, dar permiso, tomar represalia, tomar nota, tomar conciencia, tomar iniciativa, tomar precaución, tomar acuerdo, tomar distancia, tomar determinación, tomar descanso, tomar baño, hacer carrera, hacer énfasis, hacer broma, hacer colección, hacer esfuerzo, hacer referencia, hacer uso, hacer ejercicio, hacer análisis, hacer investigación, hacer declaración.

b. give advice, give example, give notice, give chase, take advantage, take action, take breath, take risk, make progress, make profit, make adjustment, make mention.

---

[55] The list of LVCs here is restricted to the corpus sample used in this dissertation.

c. donar suport, donar resposta, donar consell, donar permís, donar empenta, donar conferència, donar ajuda, donar explicació, prendre impuls, prendre bany, prendre represàlia, prendre resolució, prendre consciència, prendre nota, prendre possessió, prendre partit, prendre iniciativa, prendre precaució, prendre distància, prendre determinació, fer referència, fer esforç, fer feina, fer ús, fer anàlisi, fer dibuix, fer viatge, fer visita, fer fotografia.

d. Antwort geben, Erlaubnis geben, Rat geben, Beschreibung geben, Unterricht geben, Versprechen geben, Warnung geben, Abschied nehmen, Abstand nehmen, Rücksicht nehmen, Einsicht nehmen, Einfluss nehmen, Anteil nehmen, Platz nehmen, Anstoß nehmen, Hoffnung machen, Anfang machen, Sport machen, Eindruck machen, Gedanke machen, Sorge machen, Angabe machen, Erfahrung machen.

It seems that the combinatory limitations of bare nouns are of semantic nature (see Table 64). They can combine both with light verbs and heavy verbs (of a certain type: *have*-predicates, Borthen, 2003), and their behaviour is parallel: they take narrow scope in negation (44-45) and they are number neutral (46-47); but there is a clear difference regarding their modification possibilities, as PI-nominals only accept kind modification (48) while bare nouns in LVCs are more flexible (49-51), as well as the semantic restrictions of well-established and institutionalized activities in PI-nominals (52) and a wider range of events in LVCs (53).

Properties	PI-nominals	Bare NVEs in LVCs
Object doubling	Yes	Yes
Narrow scope in negation	Yes	Yes
Number neutrality	Yes	Yes
Modification restrictions	Yes	No
Well-established activities	Yes	No

Table 64. Summary of comparison between PI-nominals and bare NVEs in LVCs

Apart from these divergences, the main division between the PI-nominals and LVCs with bare nouns lies in the weight of the verb. As Fleischhauer (2020) highlights, the verbal head of an LVC is a light verb, whereas it is a heavy verb with PI-nominals. Besides this fact, bare nouns in LVCs are close to PI-nominals in constructions with *have*-predicates, with the exception of the adjective selection and the semantics of the constructions (as summarised in Table 64). The main reason argued in the literature regarding the difference between both groups is that bare nominals in LVCs behave the

way they do because they are part of a complex predication, while PI-nominals are not (Espinal & McNally, 2011; Fleischhauer, 2020).

In order to see if the LV and the NVE are in a complex predication, several tests can be applied but they are language dependent (Butt 1995, 2010). Some of the tests used for complex predication in Germanic and Romance languages have already been used in this chapter, that is, passivization, determination and pronominalization. The impossibility to passivize, collocate with a determiner and pronominalize are proof of monoclausality and, hence, complex predication.

Mendívil-Giró (1999: 277-404) puts forward an analysis for Spanish LVCs only when the NVE is bare based on the reanalysis. In his proposal, the LV differs depending on the NVE it selects. On the one hand, there are *support verbs* with argument structure but light or no lexical meaning, which do not undergo a syntactic reanalysis and, hence, there is a regular verb-argument relation (as in *hacer una mención,* 'make a mention'). On the other hand, there are *vicar verbs* without argument structure, which undergo a syntactic reanalysis with the NVE and are, thus, in a complex predication relation with it (as in *hacer mención*, 'make mention').

In contrast, Laca (1999: 920-921) argues that the reason for the bare nominals in LVCs not to accept pronominalization and determination is related to their non-specificity, which is due to the lack of a determiner (or plurality mark). Therefore, it should not be attributed to the effect of the LV itself, but rather to the internal structure of the nominal phrase.

Other tests related to complex predication (as proposed in Matute Martínez, 2012: 959-960) show similar behaviour in the two types of LVCs: first, the possibility to be paraphrased with a synthetic verb is common for both LVCs with a determined and a bare nominal (54), although there are exceptions in both cases (55)[56]; second, the adjacency

---

[56] As mentioned in previous chapters, there are four possibilities regarding the relation between the LVC and a synthetic counterpart: (i) the synthetic counterpart is equivalent in both its morphology and semantics, (ii) there is morphological affinity but no semantic equivalence, (iii) there is only a semantic equivalence, and (iv) there is no synthetic counterpart within the language.

(i) *give a kiss – kiss*

(ii) *hacer fiesta* ('do holiday') *– festejar* ('celebrate')

(iii) *Platz nehmen* (lit. 'take (a) place') - *setzen* ('sit')

(iv) *fer esport* ('do sport')

between the LV and the NVE would be a proof of complex predication, but other elements can be inserted between the LV and the NVE in both cases (56).

(54)   a. make mention – mention

        b. tomar distancia – distanciar(se) ('to distance')

        c. donar consell – aconsellar ('to advise')

        d. Unterricht geben – unterrichten ('to teach')

(55)   a. take advantage

        b. *Abstand nehmen* ('reject')

        c. *hacer deporte* ('do sport')

        d. *fer broma* ('joke')

(56)   a. We gave also advice   (Google)

        b. Er nimmt schnell Abschied   (Google)

        'He bade quickly farewell'

        c. Els dos germans van prendre de nou distància amb el rei francès

                                                (Google)

        'The two siblings took again distance from the French king'

        d. El rey hizo entonces declaración   (Google)

        'The King made then statement'

These tests are, therefore, inconclusive and do not point towards a complex predication analysis of LVCs. One of the properties of NVEs in LVCs which is found to be specific of these constructions (in contrast to PI-nominals) is related to the adjectival modification possibilities.

For Dobrovie-Sorin et al. (2006), as well as for Espinal & McNally (2011), PI-nominals are syntactic arguments of the verbs they combine with, but not semantic arguments. Their model is restricted to *have*-predicates in combination with bare nouns, which modify the verb.

However, studies within the neoconstructionist approach based on data from Rioplatense Spanish have proposed that there is no need for a pseudo-incorporation account (Rinaldi, 2018). For Rinaldi (2018), bare nominals such as those analysed by Espinal & McNally (2011) have an argument status: they have an indefinite and non-specific reading, have a DP projection and are of type <e>.

In contrast, Oggiani (2021: 324-325) presents a formal model within the pseudo-incorporation paradigm (following Dayal, 2011; Baker, 2014; but contra Espinal & McNally, 2011) arguing that the bare nominals denote a property of an individual and project up to NumP. Syntactic tests prove a syntactic dependency of the noun with respect to the predicate since it does not allow topicalization and they have restrictions on modification, as in (57).

(57)   a. Hoy presentan disco {internacional/*interesante}

‘They present today {international / interesting} CD’

b. Terminé monografía {de Semántica/*que quiero entregar a fin de año}

‘I finished essay {of Semantics / that I want to hand in at the end of the year}

This shows that they do not behave as canonical arguments and are syntactically dependent on the predicate, which leads to a pseudo-incorporation analysis. However, the fact that they do accept type-modification suggests that they project some structure above N. Following Dayal (2011) and Baker (2014), Oggiani (2021: 326) proposes that bare nominals in Spanish undergo a process of pseudo-incorporation in which the noun builds a NumP rather than an NP, as in (58), but still sticks to the assumption that pseudo-incorporation involves nominals which do not project a DP.

(58)
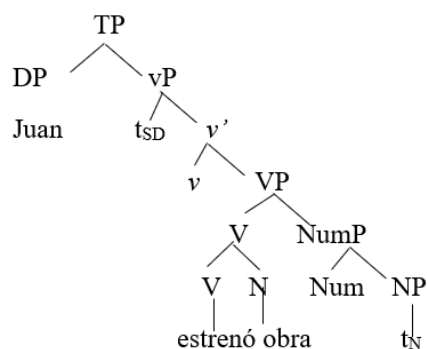
(Adapted from Oggiani, 2021: 330)

The analysis proposed consists of the movement of the bare nominal *obra* to the verbal domain, where it works as a single constituent together with the verb. This can explain the strict adjacency of these nouns and their incapability of moving to a topicalized or focused position. At the same time, the NumP projection allows the adjectival modification to appear in the nominal domain without the need to project a DP.

One of the major contributions of this analysis, as well as Baker's (2014) model for pseudo-incorporation in Tamil and Sakha, is that it allows any kind of projection under the DP (such as NumP) to account for bare nominals.

However, their pseudo-incorporation proposal cannot account for the LVC instances, where adjacency between the bare NVE and the LV is not strict, and they can insert some adjective between both elements of the construction. Moreover, it does not explain the instances with prenominal adjectives in Romance languages either (59).

(59)   a. Ha de prendre {plena/ bona/ *mala} consciència d'aquests dominis
          'S/he has to take {full/ good/ bad} awareness of these territories'
          b. S'hauria de prendre {bona/ *mala} nota de la lliçó d'honestedat
          'One should take {good/ bad} note of the honesty lesson'

In a later proposal by Oggiani (2022), she argues that there is no need for a pseudo-incorporation model to account for bare nominals in Rioplatense Spanish, such as (60), which are analysed as weak definites.

(60)   Juan mira tele / escucha radio.
          'Juan watches tv / listens radio'

Her proposal is that these bare nominals project a defective syntactic structure that contains neither D nor NumP. The fact that they are not strictly adjacent to the verb (61a), admit extraction in questions (61b) and can be coordinated with DPs (61c) suggests that they are not pseudo-incorporated structures.

(61)   a. Juan toma todas las mañanas ómnibus.   (Oggiani, 2022: 255)
          'Juan takes the bus every morning.'
          b. ¿Qué miró Ana? Solo tele.   (Oggiani, 2022: 256)
          'What did Ana watch? Only tv.'
          c. Se tomó bondi para ir y el/ un uber para volver.   (Oggiani, 2022: 257)
          'He took the bus to go and the/a uber to come back.'

In fact, the pseudo-incorporation phenomenon requires a closer relation between the predicate and the noun (Dobrovie-Sorin et al., 2006; Dayal, 2011).

With the aim to make sense of the structure of the nominal element within LVCs, the next section focuses on analysing the possibilities of determination and modification of the NVE in the LVCs under study.

6.4. Typology

As discussed so far, there are divergences between nominals introduced by a determiner and bare nouns across languages. These differences are not restricted to LVCs and can be found in other structures as well, such as *have*-predicates. In this section, the focus will be placed specifically on LVCs with a basis on the corpus data presented previously in Chapter 3. I will introduce a typology proposal for LVCs in four languages: English, German, Catalan and Spanish.

Among the syntactic and semantic properties of nominals discussed in §6.3, there seems to be a tight connection between their modification possibilities and their flexibility with determiners (as previously observed in Alonso-Ramos, 2004; Iordăchioia, 2020). That is, the presence of the article or determiner is linked to the modification possibilities in certain cases: if the noun is modified, the determiner needs to be present, as in (62). This kind of restriction is, however, limited to some specific kinds of bare singulars, as bare plurals allow a wider range of modification and present fewer or no restrictions with determiners.

(62)   Nancy hace *(una) campaña intensa contra el tabaco.

(Alonso-Ramos, 2004: 198)

'Nancy does an intense campaign against smoking'

Moreover, Alonso-Ramos (2004: 200) argues that some LVCs with bare nouns have a closer meaning to their corresponding synthetic verb (e.g. *hacer uso,* and *usar*) than when they select a noun with determination and modification (e.g. *hacer {un gran uso / un uso eficaz}*). However, this does not mean that the noun is more eventive when the determiner is absent, but rather that its absence disallows certain syntactic operations (as shown in Table 62).

For instance, in (63), there is an example of a Spanish LVC *hacer carrera* 'make career', which is generally used with a bare nominal, but it allows a qualitative adjectival modification when a determiner is introduced (63a), while the only acceptable adjective without this determiner is a relational one, *profesional* 'professional' (63b)[57].

---

[57] In this case, there are also semantic differences in the use of the LVC with a bare noun (*hacer carrera*) and with a DP (*hacer una carrera)* in Peninsular Spanish, where the latter has the specific meaning of studying a university degree. Such specification in the meaning of certain LVCs is in line with the findings in the correlation between LVCs and their synthetic counterpart (presented for Catalan in Chapter 4).

(63)   a. que había hecho una carrera tan brillante en el exterior.

<div align="right">(CORPES XXI, 2005)</div>

'who had done an outstanding career abroad'

b. que había hecho carrera profesional {*brillante}

'who had done professional career'

Having detected a correlation between determination and adjectival modification in LVCs, the following typology is proposed: LVCs with a determiner and flexible modification (64); LVCs with bare deverbal nominals, which accept both qualitative adjectival modification (65a) and relational adjectives (65b); and LVCs with non-deverbal nominals, which accept only relational modification (66a) but reject qualitative modification (66b). Additionally, an exceptional case of noun incorporation in the Rioplatense variety of Spanish (67) is also presented.

(64)   Sometimes you've got to *take a deep breath*.   (COCA, 2008)

(65)   a. no es posible *dar respuesta precisa* aún a ciertas preguntas.

<div align="right">(CORPES XXI, 2001)</div>

'it is not possible to give precise answer yet to certain questions'

b. També es *pren consciència institucional*   (CTILC, 2011)

'Institutional awareness is taken too'

(66)   a. Ich habe gar nicht extra Sport gemacht.   (DWDS, 2017)

'I didn't make any extra sport'

b. Ich habe gar nicht *(einen) {richtigen / gefährlichen} Sport gemacht

'I didn't do any {right/ dangerous} sport'

(67)   Gloria se hizo la desentendida, dio vuelta la cara (CORPES XXI, 2001)

'Gloria pretended not to know, she turned her face'

The theoretical ground of this typological classification lies in two models. First, the neoconstructionist approach to argument structure developed in Acedo-Matellán (2016) and Myler (2016). Second, the compositional hierarchy within the DP, where the structure above N determines the integration degree of the nominal with the verb. Following Grimshaw (1991, 2005), the notion of extended projections provides a syntactic approach to the nominal domain: it refers to a maximal projection which projects from a category which it shares categorial features with.

In the next sections, the three types of LVCs identified on the grounds of the corpus-based study will be presented: type 1 includes LVCs with a determiner and flexible modification in §6.4.1, type 2 involves LVCs with bare nominals in §6.4.2, and finally type 3 describes an exceptional case of incorporation with one LVC used in Rioplatense Spanish in §6.4.3.

6.4.1. Type 1: LVCs with a determiner and flexible modification

The first type of LVCs proposed are LVCs with a determiner introducing the nominal, which optionally allow the presence of modification. As discussed in Chapter 3, the determiner class in this dissertation includes both determiners and quantifiers[58].

When the nominal is introduced by a determiner (or quantifier), modification is flexible. More specifically, when the noun is deverbal, it can accept even several adjectives (68).

(68)   a. Le dio un largo beso en la boca a su marido. (CORPES XXI, 2001)
        'She gave a long kiss in the mouth to her husband'
        b. la nau que va fer el primer viatge tripulat a la Lluna. (CTILC, 2011)
        'the spaceship that made the first manned trip to the Moon'
        c. I will do my best to give brief and unbiased advice   (COCA, 2011)
        d. Nur einmal wollte ihr ein Mann einen sogenannten guten Rat geben.
                                                (DWDS, 2011)
        'Only once a man wanted to give her a so-called good advice'

In the case of non-deverbal nouns, adjectival modification is facilitated by the presence of the determiner (69). That is, when the nominal is bare, modification is not possible; whereas the presence of a determiner allows it without restrictions (as noticed by Alonso-Ramos, 2004: 198).

(69)   a. no siempre hacían un ejercicio prolijo o riguroso. (CORPES XXI, 2001)
        'they did not always do a thorough and rigorous exercise'
        b. la antropologia definia la cultura prenent un clar partit per l'home.

---

[58] The classes analysed as determiners in our database for the four languages: articles, demonstratives, possessives and quantifiers. Although in some generative approaches, they are analysed as belonging to different layers within the nominal domain: for instance, "an article is considered to be the head of the DP projection, whereas quantifiers are analyzed as heads or specifiers of a hierarchically higher Quantifier Phrase QP, accounting this way for their possible cooccurrence, as in allQ theD students" (Gianollo et al., 2022: 4).

(CTILC, 2004)

'the anthropology defined culture taking clear sides with the mankind'

c. wir hatten uns draußen einen schönen Platz genommen   (Google)

'we had taken us a good seat outside'

There are no examples of English LVCs which follow this pattern in our sample, since the majority of nouns are deverbal and the non-deverbal cases do allow modification without the need to be introduced by a determiner, as in (70).

(70)   The first companies to take full advantage of these new developments, unsurprisingly, were Google and Facebook.   (COCA, 2019)

Besides the ease in modification, the NVEs introduced by a determiner show argumental properties and pass the corresponding tests, such as the possibility to passivize (71) and pronominalize (72), among others (see Table 62).

(71)   a. La administradora hizo una aclaración.
        'The administrator made a clarification'
        b. La aclaración fue hecha por la administradora del FGD
                                                    (CORPES XXI, 2002)
        'The clarification was done by the FGD administrator'
(72)   a. Havia pres una decisió   (CTILC, 2011)
        'S/he had taken a decision'
        b. Havia pres una decisió i la va comunicar en roda de premsa.
        'S/he had taken a decision and announced it in press conference'

My proposal of analysis assumes Myler (2016) and Acedo-Matellán (2016) in that the complement of the LV is a DP, as exemplified previously in (39), which is selected by *v* in a transitive construction (or ditransitive in some of the *give*-LVCs).

In fact, when all subclasses of modification are considered (adjectival modification, PP, genitive and relative clauses), the corpus data shows that modification in determined nominals is predominant only in English (58.99%). The rest of the languages are less present: Catalan (39.58%), Spanish (37.50%), and German (25.12%) to a lesser extent (see Figure 56). The differences between languages are statistically significant ($\chi 2(3)= 847.46$, $p<.0001$, Cramer's V = 0.249).
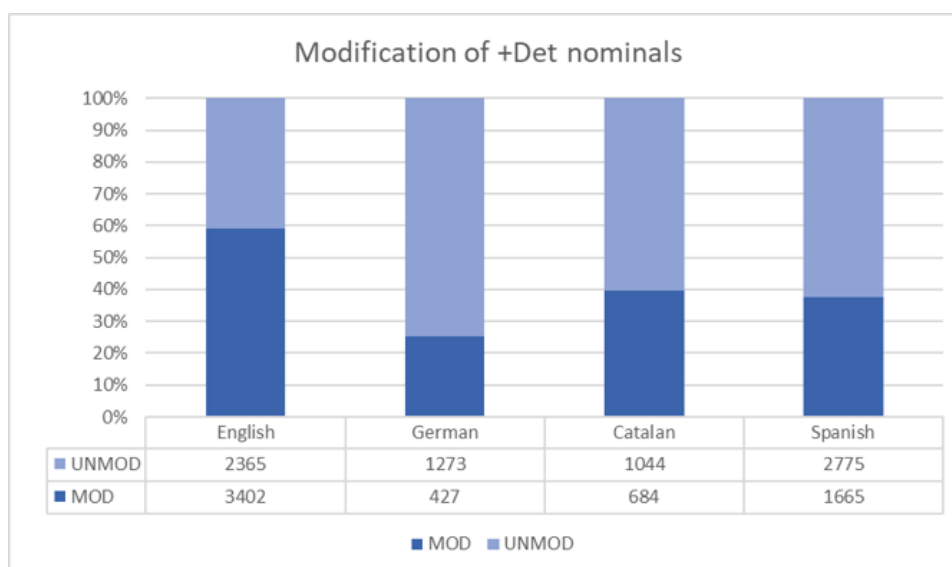
Figure 56. Modification of nominals introduced by a determiner (own data)

When all classes of modification are included (see Table 65), the most predominant class of modification is adjectival modification: 42.92% in English, 22.35% in German, 24.42% in Catalan and 21.94% in Spanish. In the case of the two Romance languages, PP modification in Catalan and Spanish is quite present (9.61% and 11.78%, respectively), while it is less relevant in the two Germanic language (5.17% in English and 0.24% in German). The rest of subclasses are residual in all languages, except for noun classifiers in English (5.15%).

Modification	English	German	Catalan	Spanish
Unmod (No)	2365 (41.01%)	1273 (74.88%)	1044 (60.42%)	2775 (62.50%)
Adj	2475 (42.92%)	380 (22.35%)	422 (24.42%)	974 (21.94%)
PP	298 (5.17%)	4 (0.24%)	166 (9.61%)	523 (11.78%)
Rel	153 (2.65%)	18 (1.06%)	29 (1.67%)	67 (1.51%)
NC	297 (5.15%)	1 (0.06%)	-	-
Gen	11 (0.19%)	11 (0.65%)	-	-
Various	168 (2.91%)	13 (0.76%)	67 (3.88%)	101 (2.27%)
TOTAL	5767	1700	1728	4440

Table 65. Distribution of modification subclasses of nominal introduced by a determiner in LVCs (all four languages)

After all, these results follow the general tendencies of the languages regarding the presence of modification in the three verbs analysed, as presented in depth in Chapter 3.

6.4.2. Type 2: LVCs with bare nominals

As predicted in Longobardi (2005), among others, bare nominals in LVCs cannot behave as canonical arguments due to the lack of the determiner: they cannot be passivized (73b) and pronominalization is restricted to the partitive, in languages where this pronoun is in use (i.e. Catalan, 73c).

(73)   a. El Consell va prendre nota de les propostes alternatives.

    'The Council took note of alternative proposals'

    b. *Nota va ser presa (pel Consell).

    'Note was taken by the Council'

    c. Vam prendre nota de la conferència, però en/#la vam prendre a mà.

    'We took note of the conference, but we took it by hand'

Therefore, they cannot take a DP as the complement, but rather some other projection below D (as proposed in Oggiani, 2021, 2022). They also present some restrictions related to adjectival modification, which will be presented in the following types of LVCs.

When the LV selects a bare nominal, there are certain limitations to modification, especially with adjectival modification. These restrictions have proven different in deverbal nominals and in non-deverbal nominals.

When the deverbal nominal is bare, the modification is more common with relational adjectives, as in (74). These relational adjectives in Romance languages are defined as to establish relations between the entities and the kind expressed by the noun (Fábregas, 2017).

(74)   a. no se hará uso legal de ella si no es estrictamente necesario.

                                 (CORPES XXI, 2002)

    'no legal use will be made of it if it is not strictly necessary'

    b. una entitat catòlica que donava suport mèdic i financer a les dones.

                                   (CTILC, 2006)

    'a Catholic entity that gave medical and financial support to the women'

    c. Finally, this Comment will urge the federal government to take legislative or regulatory action.   (COCA, 2018)

    d. Politischen Einfluss nimmt die Musik heute nicht mehr?

                                   (DWDS, 2001)

'Doesn't the music get political influence anymore today?'

However, the corpus data also provides instances of bare nominals in combination with qualitative adjectives, as in (75), which goes against the properties of bare nouns in *have*-predicates as in Stvan (2009) for English, and Espinal & McNally (2011) for Catalan and Spanish. The restrictions they posed for bare nouns in object position in the selection of kind modification do not hold for bare NVEs in LVCs. In (75), the bare NVEs are modified with qualitative adjectives, such as Sp. *plena 'full'* (75a), Cat. *eficient* 'efficient' (75b), Eng. *clear and decisive* (75c) and Ger. *besondere* 'great' (75d).

(75)   a. de los que deben *tomar conciencia plena* nuestra juventud.

                                        (CORPES XXI, 2001)

      'of which they must become fully aware of our youth'

      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.   (CTILC, 2001)

      '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'

      c. the realization that we need to *take clear and decisive action* to ensure the viability of our Catholic schools.   (COCA, 2008)

      d. "Auf Familien wird durch Freibeträge *besondere Rücksicht* genommen", teilte das Ministerium mit.   (DWDS, 2003)

      'Families are taken into great consideration through tax allowance, informed the Ministry'

As observed in Chapter 3, there is a narrower variety of adjectives which combine with LVC accepting bare nominals, as they are most often combined with relational adjectives, exemplified in (76).

(76)   a. una entitat catòlica que donava *suport mèdic i financer* a les dones

                                          (CTILC, 2006)

      'a catholic institution which gave medical and financial support to women'

      b. i *dóna suport tècnic i personal* a la resta d'unitats. (CTILC, 2010)

      'and [it] gives technical and personal support to the rest of the units.'

      c. catalans i balears es donaran *suport mutu*.   (CTILC, 2005)

      'Catalan and Balearic [people] will give mutual support'

They allow a wider range of adjectives when the bare nominal is plural (77), where both relational adjectives (77a) and qualitative adjectives (77b) are found. This difference can also be traced back to the proposal of Cyrino & Espinal (2019) that bare plurals have a D layer in their syntax which is introduced by the plural marker.

(77)   a. cosa que permetria *donar respostes educatives* als alumnes menys motivats pels ensenyaments més teòrics   (CTILC, 2000)
'which would allow to give educative answers to the students least motivated by the more theoretical topics'
b. No tenia forces per donar *respostes més llargues*. (CTILC, 2001)
'(He/She) did not have energy to give longer answers'
c. Els càlculs més fiables realitzats per experts donen *una resposta contundent* als dubtes sobre el perill d'una guerra de destrucció massiva.
                                                                    (CTILC, 2000)
'The more reliable calculations made by experts give a conclusive answer to the doubts regarding the doubts about the danger of a mass destruction war'

In absolute numbers (Figure 57), corpus data support a higher presence of modification in bare plurals in three languages: English (66.34% of modified bare plurals vs. 60.18% bare singulars), German (38.35% of modified bare plurals vs. 8.47% bare singulars), and Catalan (35.78% of modified bare plurals vs. 18.46% bare singulars). Spanish LVCs show a different tendency: bare singulars are modified slightly more often (31.71%) than bare plurals (27.27%). It is important to note that all subclasses of modification are included here, it is not restricted only to adjectival modification, and Spanish tends to select PP modification more commonly than the rest of the languages, especially than the Germanic languages (as seen previously in Table 65).
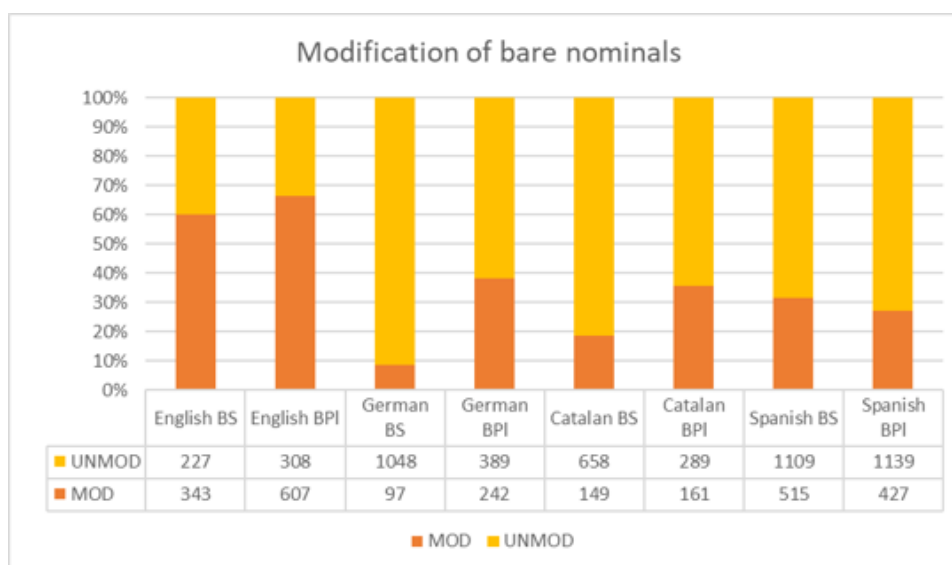
Figure 57. Modification of bare nominals: bare singulars (BS) and bare plurals (BPl)

As observed in Chapter 3, when all classes of modification are included in the modification of bare singulars (Table 66), adjectival modification is prevalent in the two Germanic languages (24.56% in English and 8.30% in German); while PP modification is more present in the two Romance languages (11.40% PP modification versus 6.07% adjectival modification in Catalan, and 23.95% PP modification versus 5.23% adjectival modification in Spanish). Therefore, the presence of PP modification in the two Romance languages increases the overall results of modification in bare singulars (shown in Figure 57).

Modification	English	German	Catalan	Spanish
Unmod (No)	227 (39.82%)	1048 (91.53%)	658 (81.54%)	1109 (68.29%)
Adj	140 (24.56%)	95 (8.30%)	49 (6.07%)	85 (5.23%)
PP	52 (9.12%)	2 (0.17%)	92 (11.40%)	389 (23.95%)
Rel	3 (0.53%)	-	2 (0.25%)	7 (0.43%)
NC	9 (1.58%)	-	-	-
Gen	26 (4.56%)	-	-	-
Various	113 (19.83%)	-	6 (0.74%)	34 (2.10%)
TOTAL	570	1145	807	1624

Table 66. Distribution of modification subclasses of bare singulars in LVCs (all four languages)

In contrast, when all classes of modification are included in the modification of bare plurals (Table 67), the presence of PP modification does not seem to have had an influence on the overall modification results. In this group, all languages have a greater prevalence of adjectival modification above the rest of the classes: 51.91% in English, 35.50% in German, 23.78% in Catalan and 15.52% in Spanish. These are, also, always

higher than adjectival modification of bare singulars in each language. Also, it is important to highlight that PP modification is still more common in Romance languages also for bare plurals, although to a lesser extent than for bare singulars (see Table 66): 9.78% of bare plurals in Catalan LVCs is modified by a PP, and 10.34% for Spanish.

Modification	English	German	Catalan	Spanish
Unmod (No)	308 (33.66%)	389 (61.65%)	289 (64.22%)	1139 (72.73%)
Adj	475 (51.91%)	224 (35.50%)	107 (23.78%)	243 (15.52%)
PP	28 (3.06%)	4 (0.63%)	44 (9.78%)	162 (10.34%)
Rel	14 (1.53%)	10 (1.58%)	6 (1.33%)	12 (0.77%)
NC	59 (6.45%)	-	-	-
Gen	1 (0.11%)	1 (0.16%)	-	-
Various	30 (3.28%)	3 (0.48%)	4 (0.89%)	10 (0.64%)
TOTAL	915	631	450	1566

Table 67. Distribution of modification subclasses of bare plurals (all four languages)

In sum, bare nominals in LVCs accept modification, which is not restricted to kind modification (as proposed for bare nouns in *have*-predicates in Espinal & McNally, 2011). However, some differences between bare plurals and bare singulars in LVCs have been detected: while bare plurals show no restrictions with regard to modification and they are more often modified, bare singulars present a tendency to be combined with relational adjectives more often than qualitative adjectives (although these are not completely ruled out) and they are generally less often modified than bare plurals.

After all, both kinds of bare nominals accept all classes of modification (see Tables 66 and 67) although adjectival modification is predominant in Germanic languages both in the case of bare singulars and bare plurals, and also in Romance languages with bare plurals. In contrast, bare singulars in the LVCs of the two Romance languages show a tendency to select PP modification, as they can select different kinds of PPs as modifiers (e.g. *hacer uso de* 'make use of' selecting quasi-argumental PPs, and not only attributive PP modification, as discussed in Chapter 3).

The two kinds of adjectives (qualitative and relational) have different positions within the nominal domain, and the theoretical analysis should account for a more internal position of the relational adjectives than that of qualitative, which have more mobility within the phrase (can appear both in prenominal and postnominal position). As Fábregas (2017: 75) points out, the position within the nominal domain of relational adjectives taking into account their restrictions (that is, they must be postnominal and as internal as

possible to the head noun) has not reached a consensus – although it is clear that they must occupy different positions (as pointed out in Bonet & Solà, 1986, for Catalan; and Bosque & Picallo, 1996, for Spanish).

Fábregas (2017) proposes an analysis within the extended projection framework of an NP with three layers: nP, NP, and the root (78). Given this structure, relational adjectives are in the specifier of NP in all languages. However, the head N moves up to n allowing the postnominal position in languages like Spanish and Catalan.

(78)
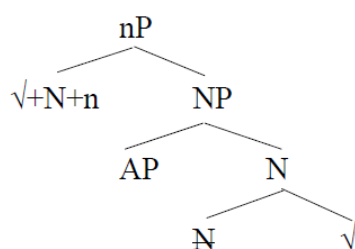
(Fábregas, 2017: 84)

Partly following Oggiani's (2021, 2022) analysis of bare nouns in Rioplatense Spanish (discussed previously in §6.3), bare NVEs cannot project D, but they can project up to NumP because they can select adjectives.

Regarding the connection between the LV and the bare nominal, the latter cannot be analysed as canonical arguments of the verbs (see Table 62 for argumenthood tests); but they accept elements to be inserted between the LV and the NVE, as in (79). For this reason, a pseudo-incorporation analysis does not hold.

(79) a. Tomaron *de nuevo* conciencia {social / precisa/ plena} de los problemas
'They got again {social/ precise/ full} awareness of the problems'
b. Van donar *per fi* resposta {afirmativa / clara / visible} a la gran pregunta
'They finally gave {positive/ clear/ visible} answer to the big question'
c. They made *finally* {professional/ interesting / silly} conversation
d. Sie nehmen *gleichzeitig* {literarischen/ großen / nennenswerten} Einfluss
'They take at the same time {literary/ great/ considerable} influence'

In contrast, the non-deverbal bare nominals in LVCs exhibit some differences regarding modification.

On the one hand, all four languages allow modification by means of an indefinite quantifier, as Sp. *mucho* and Cat. *molt* 'a lot', as in (80a,b); and also with Eng. *a lot* and Ger. *viel* 'a lot' (80c,d).

(80)   a. En los dos últimos meses he hecho *mucho* ejercicio

'In the last two months, I have done a lot of exercise'

b. L'any passat va donar *molt de* consell

'Last year, s/he gave a lot of advice'

c. While I was able to take *a lot more* advantage of the situation.

(COCA, 2012)

d. die im Urlaub *viel* Sport gemacht haben.   (DWDS, 2013)

'who have done a lot of exercise on holidays'

On the other hand, when the bare nominal is non-deverbal, there is a clear divergence between Romance and Germanic languages in the empirical sample. As in (81), Spanish and Catalan instances can accept the modification by a relational adjective, like Sp. *aeróbico* 'aerobic' and Cat. *mèdic* 'medical' but reject qualitative modification with adjectives like Sp. *extenuante* 'exhausting' or Cat. *interessant* 'interesting'.

(81)   a. Almudena Cid hace ejercicio {aeróbico / gimnástico / *extenuante}

'Almudena Cid does {aerobic/ gymnastics/ exhausting} exercise'

b. Va donar consell {mèdic / formal / *interessant}

'S/he gave {medical / formal / interesting} advice'

In contrast, Germanic languages allow qualitative modification with non-deverbal bare nominals (82). Hence, they do not differ from deverbal nouns; their prenominal position does not pose a need for N movement, as they do in Romance languages.

(82)   a. you had more money but perhaps less ability to take *full* advantage of it.

(COCA, 2019)

b. Es geht nicht darum, ob du *sauberen* Sport machst. (DWDS, 2008)

'It's not about whether you do clean exercise'

The fact that these non-deverbal nouns do not accept a postnominal qualitative adjective in Romance languages proves that they cannot undergo N movement as deverbal nouns in (78). Although this movement is not possible, they do not show proof

of any kind of incorporation since there is no necessary adjacency between the LV and the bare non-deverbal nominal: adverbials can be introduced, as in (83).

> (83)   a. En los dos últimos meses he hecho *raramente* ejercicio.
>
>   'In the last two months, I have done rarely exercise'
>
>   b. Li cal donar *de nou* consell als estudiants.
>
>   'S/he must give again advice to the students'

The proposal of analysis for bare nominals in LVCs supports Oggiani's analysis of bare nouns (2021, 2022) in that the complement of the LV is a NumP, which is selected by *v*, as exemplified previously in (58). Moreover, it is important to bear in mind that the adjectives are inserted in different levels within the NumP, depending on their closeness to the nominal (as in Fábregas, 2017).

6.4.3. Type 3: An exceptional case of incorporation

There is an exceptional case of an LVC that does not fall into any of the previous classes and has a very restricted use. It is the case of *dar vuelta* in Rioplatense Spanish (lit. 'give turn')[59] which can be built with a direct object, *la cabeza* ('the head'), as in (84a), and can also be pronominalized with the accusative pronoun *la* (84b).

> (84)   a. Para que un día un macho cualquiera te diera vuelta la cabeza como una
>
>   veleta.   (CORPES XXI, 2001)
>
>   'So that one day any man would turn your head around like a weathercock'
>
>   b. te *la* diera vuelta como una veleta.
>
>   'would turn it around like a weathercock'

It is important to note that this is not a case of accusative doubling (typical from Rioplatense Spanish, as analysed in Zdrojewski & Sánchez, 2014; Di Tullio et al. 2019) because it can also pronominalise a masculine antecedent with *lo,* as in (85b), which refers to *el partido* ('the match'), as in (85a).

> (85)   a. De la mano de Messi, Barcelona dio vuelta el partido.   (Google)
>
>   'Thanks to Messi, Barcelona turned the match'

---

[59] This use differs from the LVCs with a DP (una vuelta) as in (i) which means to go around something, as well as from its use with the plural noun *dar vueltas* (ii) which has a more collocate meaning of thinking or overthinking. These two uses are common in all Spanish varieties.
    (i) Dio una vuelta a la casa.
    'He walked around the house'
    (ii) Estuvo dándole vueltas toda la noche.
    'He was overthinking it the whole night'

b. Con uno menos, Estudiantes se *lo* dio vuelta a River.   (Google)

'With one less, Estudiantes turned it to River'

Also, the NVE *vuelta* cannot be pronominalized (86b) nor extracted in interrogatives (86c), showing that it is not the canonical argument of the LV (as pointed out by Galbarini, 2017), contrary to the direct object in (84b, 85b).

(86)   a. Juan dio vuelta la hoja = Juan volteó / giró la hoja.

'Juan turned the page'

b. *Juan la dio la hoja.

'Juan turned it the page'

c. *¿Qué dio Juan la hoja?   (Galbarini, 2017)

'What did Juan give the page?'

A third key element is the fact that there is no possibility of inserting an element between the LV and the NVE. In (87) the adverbials between these two elements are not acceptable by native speakers of the Rioplatense variety of Spanish.

(87)   Dio {*ya/ *súbitamente} vuelta el partido

Gave.3SG {already / suddenly} turn the match

'S/he gave {already / suddenly} a turn to the match'

For Galbarini (2017), *dar vuelta* is an instance of a complex predicate (in line with Borer's 2005 proposal for resultative constructions), although her analysis does not portray how the incorporation of *vuelta* to the verb takes place, as in (88).

(88)
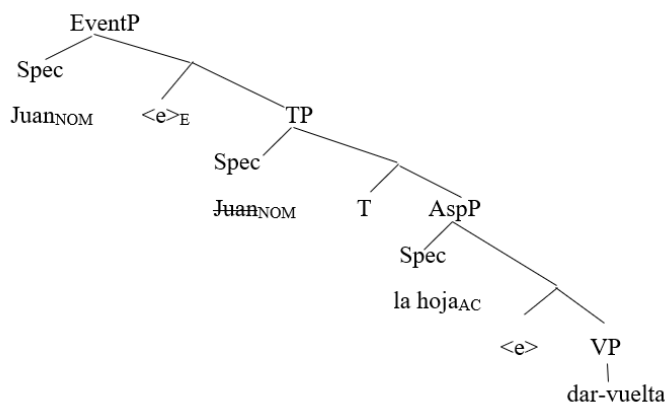
(Adapted from Galbarini, 2017: 14)

According to Acedo-Matellán & Pineda (2019), in similar instances in Basque which allow a DO, as in (89), "a lexicalization process has taken place, so that the argument structure of the LVC includes a slot for a DO with a Theme-role" (p. 206). The reanalysis in these instances from Basque LVCs is proven also in the phonology, since both elements are pronounced as single words with one accent [hutségin, hotségin]. Therefore, 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 Martínez, 2015).

(89)　a. Ume-a-k　　eskol-ak　　　huts　　egin ditu.
　　　　kid-det.abs-erg class-abs.pl failure/empty　　do aux.prs.erg3sg.abs3pl
　　　　'The kid missed the classes'
　　　　b. Gizon-a-k　berrion-a　　　　　　hots　　　egin　　　du
　　　　man-det.abs-erg good_news-det.abs noise/rumor do aux.prs.erg3sg.abs3sg
　　　　'The man spread the good news'

These properties are aligned with a case-manipulating type of incorporation (Mithun, 1984): the case left by the incorporated argument (the NVE, in this case) is available for the other argument available (the oblique argument), which is promoted. For Mithun (1984: 856), the noun loses the syntactic status as an argument of the clause, because it is incorporated by the verb, which then leaves a case position vacated that can be occupied by the oblique argument.

Going back to *dar vuelta*, in an early description of this construction, Masullo (1996) suggests that the NVE *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, since there is a ditransitive construction with two objects: one with a defective direct object (*vuelta*) and a full direct object (*hoja*).

Although double accusative constructions are not typical of present-day Spanish, they are attested in Medieval periods of the Romance languages[60], and they are common

---

[60] Matute Martínez (2012, p. 960) studies these kind of 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. The data shows that some LVCs allow this structure in Old Spanish (as in *aver mester* 'to need' and *dar vuelta* 'to turn') but disallows it with the same verbs in combination with other nouns (*aver envidia* 'to envy', *dar ayuda* 'to help'), but no semantic pattern is found (p. 965). It coexists with double accusative constructions in medieval Spanish inherited from Latin verbs (*doceo* 'to teach', *rogo* 'to beg') which are kept in ditransitive romance predicates (*enseñar* 'to teach', *rogar* 'to beg'), as well as other Romance languages such as Italian

in other language families (i.e. Persian). Their main particularity is that there is no univocal correspondence between the morphological case and the syntactic function.

As stated before, this case is exceptional in present-day Spanish and it is limited to a variety of the language spoken in el Río de la Plata (the area including Buenos Aires, Argentina, and Montevideo, Uruguay) and has a wide use of singular bare nominals that have not developed in other varieties of the language, such as the examples in (57) with *presentar disco* ('present album') and *terminar monografía* (lit. 'finish dissertation') (Oggiani, 2021, 2022). This exceptional case of *dar vuelta* followed by a direct object in Rioplatense Spanish is the only instance found cross-linguistically of incorporation of the NVE.

Interestingly, Pineda (2014, 2015) defends that some Catalan and Spanish ditransitive constructions should be analysed as double object constructions. In syntax, since Pylkkänen (2008), the applicative head is considered to be responsible for the dative codification of the IO, while the argument in the complement position (the theme DO) has to move to the Spec of v to be assigned the accusative case. Pylkkänen (2008) distinguishes two classes of applicatives: high and low applicatives. High applicatives are generated outside the VP and they relate the VP with an event-related applied argument (i.e. beneficiary) (90a); while low applicatives are generated within the VP and they relate the theme argument and the verb to an applied argument (i.e. recipient or source) (90b). For Pylkkänen (2008), low applicatives (90b) involve a dynamic transfer of possession of the theme to or from the applied argument, and she applied this analysis to English double object constructions.

(*poner mente*, *tenere mente* 'to bear in mind'). In these structures, the presence of two direct objects involves a selection of different thematic roles: the NVE is the theme, and the second object is the goal or beneficiary (p. 967). Most of these constructions are later documented as ditransitives with a DO with accusative case and an OI with dative case, or even with a PP as a second argument (*aver vergüenza de* 'to have shame, to feel ashamed', *poner por/de nombre* 'to put of name, to name/call sb'). Other combinations disappeared in favour of their synthetic counterparts (*aver mester> necesitar* 'to need', *parar mientes >atender* 'to pay attention', etc.).
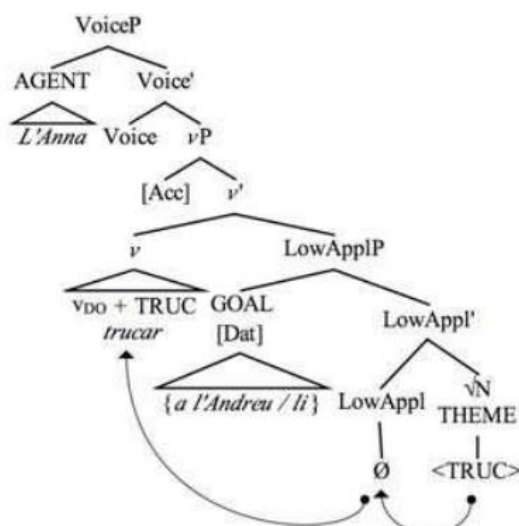
(90)
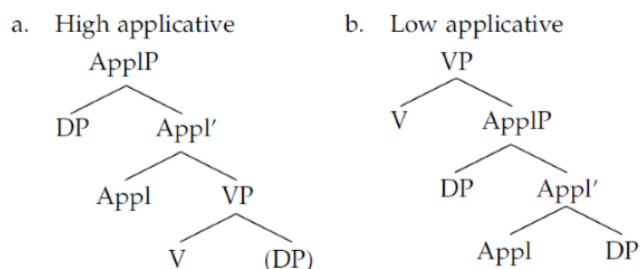
(McGinnis, 2008: 4)

This difference is relevant for Pineda's (2015) analysis of unergative verbs in Catalan, such as *trucar* ('to phone'), the DO does not reach the Spec of v and, instead, merges with the verb as in (91); while the IO (*a l'Andreu, li*) receives the dative marking through the Appl head.

(91)

(Pineda, 2015: 89)

What is crucial in Pineda's proposal is that low Appl does not behave the same across languages, since their case marking properties might change (as pointed out by Pylkkänen 2008, Cuervo 2003). In (91), a Romance applicative assigns dative case to the specifier, whereas an Anglo-Saxon applicative does not have the dative case available and it assigns accusative case by default, as in (92).

(92)   Anna gave Andreu a rose

The fact that there is a possibility in Catalan to use the accusative clitic for the IO, as in (93), does not prove that the unergative verb behaves as a transitive, but rather that

the complement has received an unexpected accusative marking. Following Bilous (2011), Pineda (2014, 2015, 2020) describes this phenomenon as a differential indirect object marking, which implies that the object is an IO although it can receive accusative marking. The semantic role of this object is that of a goal: either receiver (93a) or beneficiary (93b).

(93)   a. *telefonar, escriure, respondre* + IO

        'phone, write, answer'

        b. *ajudar, robar* + IO

        'help, steal/rob'

Consequently, the syntactic structure has to provide a slot for this complement. For Pineda (2015), the object with accusative marking is a true IO and occupies its original position in the specifier of Appl (94).

(94)
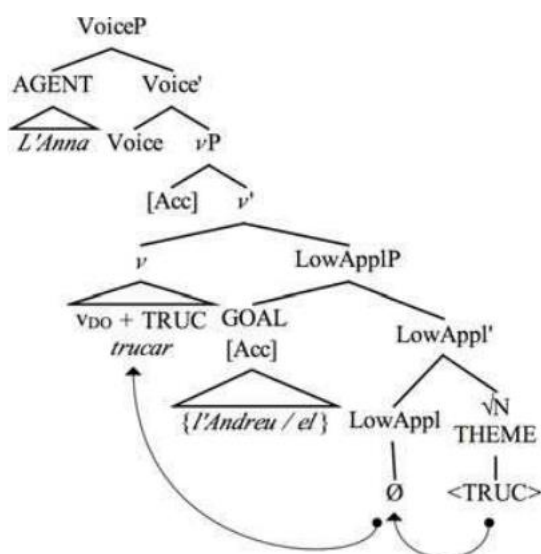
(Pineda, 2015: 90)

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, as in (91), which does not assign dative case and, hence, assigns the only available case (accusative) by default.

Passivization can be used to test that these IO have indeed received the accusative case (95), as in English double object constructions (96).

(95) a. Un brètol ha robat el Joan.

'A vandal has robbed Joan'

b. El Joan ha estat robat per un brètol.   (Pineda, 2015: 92)

'Joan has been robbed by a vandal'

(96) a. Andreu gave Anna a rose.

b. Anna was given a rose.

A parallel could be drawn with the accusative case assignation of the second object in *dar vuelta* in Rioplatense Spanish, which can also be passivized with the morphological passive (97a), and more common with the *se*-passive (97b).

(97) a. El partido fue dado vuelta y Atlético Rafaela se impuso por 2-1

(Google)

'The match was turned and Atlético Rafaela won 2-1'

b. Con solidaridad se dio vuelta el partido   (Google)

'With support the match was turned'

Pineda's analysis is also compatible with the Dependent Case theory (Marantz, 1991; Wood, 2011), which is formulated as follows:

(98) Direct Case Assignment

If a DP *a* has no inherent case feature at spell out, it is assigned accusative iff there is some other DP *b* visible to *a* where (i) *b* has no inherent case feature and (ii) *b* c-commands *a*. Otherwise, *a* will be nominative.   (Myler, 2013: 190)

This would explain why the accusative case is assigned to the oblique argument (*el partido* 'the match', in 85), since it only requires that it is c-commanded at Spell Out by a nominal that lacks inherent case.

Besides the assignation of accusative case to the oblique object, the analysis of this kind of LVCs should account for the fact that the NVE remains in situ (*vuelta*) and does not incorporate as an unergative verbs. At the same time, it should explain the fact that there is no possibility to introduce an element between the LV and the NVE (87, repeated here as 99).

(99) Dio {*ya/ *súbitamente} vuelta el partido

Gave.3SG {already / suddenly} turn the match

'S/he gave {already / suddenly} a turn to the match'

322

Within an extended projection framework (Grimshaw, 1991), the fact that the NVE does not receive the accusative marking can be explained through the internal structure within the nominal domain: *vuelta* does not project over *n*, and it cannot be case marked. The other object (oblique IO) has the accusative case available (as in the dependent case theory by Marantz, 1991) and can take it through the Applicative head proposed by Pineda (2014, 2015) in (93).

Spanish is a language that *a priori* does not fall into the group of polysynthetic languages that undergo nominal incorporation (à la Mithun, 1984; Baker, 1988), which is a morphosyntactic process where a noun integrates into a verb or preposition that selects it and creates a complex predicate. The incorporated noun is highly defective and does not have number, case or definiteness marking. 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 seen in Type 2 in our classification in (83). However, this is not the case for *dar vuelta* + DO and, hence, an incorporation analysis should be applied here to explain this unique LVC.

The fact that such an option does not exist in other LVCs, nor the rest of languages (English, German and Catalan), and that *dar vuelta* followed by a DO is restricted to one variety of Spanish (Rioplatense Spanish) shows that incorporation is not productive in LVCs in these languages.

6.5. Interim summary

The data from Germanic and Romance languages presented in this study confirm the continuum in the degree of cohesion between the light verb and the NVE (Acedo-Matellán & Pineda, 2019), where the presence of the determiner introducing the NVE is crucial. In this regard, there are some differences depending on the language. In English, a majority of LVCs present an obligatory determiner, that is, LVCs in English are mainly transitive-like. In contrast, less than half of Spanish, Catalan and German LVCs have an obligatory determiner, which means that the majority of LVCs can occur with a bare noun.

When the correlation between determination and modification is analysed, languages show different tendencies within language families. As seen in Table 68, NVEs introduced by a determiner are modified more often than bare nominals in English (58.99%), Catalan (39.58%) and Spanish (37.50%); while in German they are less often modified (25.12%) than bare plurals (38.35%). With regard to bare singulars, these are

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 6. A typology of LVCs in Germanic and Romance languages

less often modified than bare plurals in English (60.18% bare singulars versus 66.34% bare plurals) and in Catalan (18.46% bare singulars versus 35.78% bare plurals); but the tendency is reversed in Spanish (31.71% bare singulars versus 27.27% bare plurals) and in German (8.47% bare singulars and 38.35% bare plurals).

	+Det	Bare Sing	Bare Pl
English	3402 (58.99%)	343 (60.18%)	607 (66.34%)
German	427 (25.12%)	97 (8.47%)	242 (38.35%)
Catalan	684 (39.58%)	149 (18.46%)	161 (35.78%)
Spanish	1665 (37.50%)	515 (31.71%)	427 (27.27%)

Table 68. Summary of correlation between determination and modification (all subclasses)

At the end of the continuum of LVCs where the degree of cohesion between the light verb and the NVE is the highest, we can find transitive-like LVCs (Type 1). The analysis for this group of LVCs within the neo-constructionist approach to argument structure is that of a transitive creation construction, where the verb is followed by a DP that behaves like an argument (Acedo-Matellán, 2016).

At the other end of the continuum, where the degree of cohesion is the lowest, we find the LVCs with bare nominals (Type 2). The analysis proposed is that they cannot project the DP layer, but they do project other layers under D when they accept adjectival modification. The inability of noun phrases to behave like canonical arguments explains why they have been traditionally analysed as complex predicates (Butt, 2010) or pseudo-incorporated nominals (Massam, 2001; Espinal & McNally, 2011). However, there is no need for such analysis because there is no actual adjacency between the LV and the NVE.

Finally, there is an exceptional case of a LVC restricted to Rioplatense Spanish where there is a noun incorporation which leaves a free spot for an extra object to be selected as the direct object of the verb (Type 3).

The main properties of the three types are summarised in Table 69.

| | 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	Eng. *give brief and unbiased advice*	Cat. *donar resposta efectiva* 'give effective answer'	Sp. *hacer ejercicio aeróbico* 'make aerobic exercise'	Sp. *dar vuelta el partido* 'turn the match around'

Table 69. Summary of the proposed typology of LVC

To conclude, the present proposal defends that all LVCs along the continuum behave like regular verbo-nominal combinations (Bruening, 2016). The differences in the degree of cohesion between the LV and the NVE are due to the nature of the NVE rather than the light verb itself, so that differences between LVCs lie in the internal structure and semantic properties of the NVE. Observed restrictions (such as the combination of non-deverbal nouns only with relational adjectives) are aligned with general restrictions of bare nominals in the languages. Structurally, the compositional hierarchy within the DP shows the combinatory possibilities of the nominal in LVCs: the more projections in the structure, the less integration of the two LVC components and, thus, the more independence of the nominal element.

UNIVERSITAT ROVIRA I VIRGILI
THE NOMINAL IN LIGHT VERB CONSTRUCTIONS: A CORPUS-BASED STUDY IN PRESENT-DAY ENGLISH, GERMAN, CATALAN AND SPANISH
Georgina Alvarez Morera

Chapter 7. Conclusions

Chapter 7. Conclusions

This dissertation reports the results of a corpus-based investigation on the properties of the nominal component in light verb constructions in two Germanic languages (English and German) and two Romance languages (Catalan and Spanish), and takes them as a basis to suggest some improvements in the theoretical analysis of these constructions. The first section of the dissertation has provided a description of the data, which was addressed from a descriptive perspective within the Corpus Linguistics methodology. The second section has focused on a theoretical proposal for a typology based on the projection of NVE in LVCs, which was framed within the neoconstructionist approach to argument structure and the application of the extended projection model within the nominal domain.

In this concluding chapter, I first outline the main findings of this dissertation with a summary of each chapter in §7.1. Afterwards, the limitations and some areas for extension are discussed in §7.2. Last, I present the general conclusion of the dissertation and its main contributions in §7.4.

7.1. Summary of the findings

After an exhaustive review of the theoretical literature on LVCs, with a thorough description of the properties of its two main components, the LV and the NVE, and the many ways in which they interact, Chapter 3 has presented the contrastive corpus-based study based on data from four different corpora: COCA, for English LVCs; DWDS, for German LVCs; CTILC, for Catalan LVCs, and CORPES XXI, for Spanish LVCs. The database for this study was based on a total of 21,334 distributed among the four languages.

The quantitative findings focused on similarities and differences between three LVs (*give, take* and *make*) and the four languages studied, particularly on the weight of the NVE: its determination and modification properties. Regarding determination, *give*-LVCs and *make*-LVCs in all languages have predominantly NVEs introduced by a determiner, while *take*-LVCs have mostly determined NVEs in English and Spanish, but not in German and Catalan, which have more bare nouns.

In terms of modification, English LVCs show the highest proportion of modified NVEs, while German has the lowest. The two Romance languages show similar patterns,

with around a third of NVEs modified in all three verbs. Regarding subclasses of modification, adjectival modification is prevalent in all languages for the three verbs, except in Catalan *fer*-LVCs and Spanish *hacer*-LVCs ('make') where PP modification is also highly present, influenced by the selection of the NVE and its transitive base verb, as in (1), where the noun Sp. *uso* 'use' selects in the LVC a PP (1a) which is the DO of its verbal base *usar* 'to use' (1b).

(1) a. Las generalistas suelen hacer uso *de una amplia variedad de hábitats*,

(CORPES XXI, 2001)

'The generalists usually make us of a wide variety of habitats'

b. Las generalistas suelen *usar una amplia variedad de hábitats*,

'The generalists usually use a wide variety of habitats'

Moreover, the frequency of LVCs in corpora shows varying effects on modification. In *give*-LVCs, high-frequency instances are more often modified in English and Spanish, while in German and Catalan, both high- and low-frequency instances show more modification than medium-frequency LVCs. For *take*-LVCs, English and German high- and medium-frequency instances have more modification, while Catalan and Spanish low-frequency instances have increased modification. In contrast, *make*-LVCs show a tendency for low-frequency lemmas to be the most often modified in all languages. All in all, the most frequent LVCs do not show signs of lexicalization, contrary to what had been established in previous studies. Only a few highly frequent LVCs tend to co-occur with specific adjectives, while the majority of LVCs combine with a wide range of adjectives and show low tendencies for co-occurring adjectives and nouns.

Chapter 4 has explored the interchangeability of Catalan LVCs and their synthetic verbal counterparts. The study confirms previous findings in different languages, showing that LVCs with bare nouns and unmodified NVEs are statistically more likely to be interchangeable. The weight of the nominal element is an essential variable: lighter nominals (e.g. bare and unmodified) are more likely to be interchangeable with synthetic verbs, as (2) in comparison to (3).

(2) a. que utilitzem per *fer dibuixos*, (CTILC, 2007)

'that we use to make drawings, writing, mathematical calculations'

b. que utilitzem per *dibuixar,*

(3) a. una cambra fosca permetia *fer dibuixos molt realistes*, (CTILC, 2011)

'a dark chamber allowed to make very realistic drawings'

b. *una cambra permetia *dibuixar realistament*

'a dark chamber allowed to draw realistically'

Moreover, the degree of frequency in the Catalan database is also found to be relevant for interchangeability. That is, medium-frequency LVCs, followed by high-frequency ones, are more likely to be interchangeable. The most common LVCs are highly interchangeable with synthetic verbs. These findings thus reject that LVCs are used more when there is no semantic correlation with the synthetic verb, or when they have developed specific lexicalized meanings that the synthetic verb lacks.

Chapter 5 has reviewed previous approaches to LVCs and concludes, in agreement with Butt & Lahiri (2013) and Myler (2016), that no separate lexical entries are needed for full verbs and light verbs. Additionally, the chapter discusses syntactic divergences between NVEs introduced by determiners and bare NVEs, which were attributed to nominal projection levels (nP for bare nouns, DP for determined NVEs) in Kornfeld's (2004) analysis of Spanish LVCs.

Chapter 6 proposed a typology of LVCs for Germanic and Romance languages, confirming the existence of a continuum of cohesion between LVs and NVEs. The presence of a determiner introducing the NVE is crucial for its argumental properties. Thus, at one end of the continuum, LVCs with a determined NVE behave like a transitive creation construction. At the other end, LVCs with bare nominals lack a DP projection, but they can project other layers under D when accepting adjectival modification. The inability of bare noun phrases to function as canonical arguments explains why they are traditionally analysed as pseudo-incorporated nominals, although no actual adjacency exists between the LV and the NVE, and such an analysis is not needed from a syntactic perspective. Some differences have been detected with regards to modification of the NVE in the two Germanic languages in comparison to the two Romance languages. When the NVE is a non-deverbal noun, English and German NVEs can be modified through a qualitative adjective, as in (4). Such qualitative modification is not acceptable in the case of Spanish and Catalan non-deverbal NVEs, as in (5), which have a more restricted selection of the kind of adjectives, allowing only relational ones, as in (6).

(4) a. you had more money but perhaps less ability to take *full* advantage of it.

(COCA, 2019)

b. Es geht nicht darum, ob du *sauberen* Sport machst   (DWDS, 2008)

'It's not about whether you do clean exercise'

(5) a. Hace ejercicio *aeróbico diario*   (CORPES XXI, 2001)

'S/he does aerobic exercise daily'

b. Va donar consell *mèdic*

'S/he gave medical advice'

(6) a. Hace ejercicio {*extenuante / *específico}   (CORPES XXI, 2001)

'S/he does exhausting / specific exercise'

b. Va donar consell {*interessant / *útil}

'S/he gave interesting / useful advice'

Finally, an exceptional case is found in Rioplatense Spanish where there is noun incorporation (*dar vuelta + DO),* thus allowing an extra object to be selected as the direct object of the verb. The typology is summarized in Table 70.

	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	Eng. *give brief and unbiased advice*	Cat. *donar resposta efectiva* 'give effective answer'	Sp. *hacer ejercicio aeróbico* 'make aerobic exercise'	Sp. *dar vuelta el partido* 'turn the match around'

Table 70. Summary of the proposed typology of LVC

7.2. Limitations and future research

Despite recent progress and numerous advantages of considering this approach, Corpus Linguistics is a field that show some constraints that researchers should take into account when analysing corpus data and drawing conclusions from their corpus-based research. First, it is essential to acknowledge that the scope and scale of corpora are restricted; not all language variations, registers, or text types are equally well-represented in extensively and consistently compiled corpora. This limitation becomes especially relevant when the study involves four different languages extracted from four different corpora. The four selected corpora (COCA, DWDS, CTILC and CORPES XXI) have

been compiled by institutions and academics with the aim to create representative databases; however, balance can only be presupposed and not completely confirmed. In other words, the results can only reflect tendencies, which cannot be generalised and must be contrasted and triangulated with the general constraints from grammar, and native speakers' judgements of well-formedness and acceptability.

Second, the selection of LVCs analysed had to be done with certain research delimitations in mind, such as the number of lemmas analysed per verb and language (maximum of 15), the time period of the 21st Century, and the degree of frequency within the corpora (high, medium and low). The final decision was, however, dependent on the researcher's choice.

Third, the lack of automatic identification of LVCs led to a manual selection of the instances included in the corpus-based study, while the tagging for the different variables was also done manually. All possible mistakes are, thus, my own. In future research, the present tagging could be compared to that of automatic tagging systems implemented for collocations.

Future research could aim at reducing the limitations of the present study by extending the scope of the study to other LVCs, to examine in more detail other individual patterns and complementation preferences. The inclusion of other common LVs (i.e. *have* or *put*) could also be helpful to see if the main findings hold across different verbs.

Regarding the coexistence of LVCs and their synthetic verbal counterparts, the analysis should be extended to the other three language databases: English, German and Spanish, in order to be able to draw a comparison with the results from the Catalan LVCs. Also, a new corpus research of the synthetic verbal counterparts could shed light to their interchangeability by LVCs. Only with these two-way studies could one obtain a full understanding of the coexistence between the analytical and synthetic forms.

Finally, a more fine-grained analysis from the perspective of theoretical linguistics could focus on the different kinds of modification in the NVE, and their relationship with the verbal base in eventive nouns. Such an analysis at the syntax-semantics interface could contribute to the theoretical explanation of the nominal domain.

7.3. General conclusion

Despite those aspects not covered in this dissertation and the future research areas into which this investigation could be directed (described in §7.3), this thesis has contributed to the field by exploring the significance of the nominal in LVCs in four different languages: English, German, Catalan and Spanish. More specifically, the determination and modification properties of the NVE have been explored on corpus data in Chapter 3, which have later served as the basis for the typology proposal and the theoretical analysis in Chapter 6. Also, an exploration of the relationship between the LVCs and their synthetic verbal counterpart has been undertaken with a focus on the data for Catalan LVCs, thus contributing novel empirical data to the literature on this topic.

Some of the corpus-based findings raise relevant issues for theoretical linguistics, especially concerning the analysis of nouns. The selection of determined NVEs and bare nouns aligns with general tendencies of the languages, suggesting LVCs behave like traditional verbo-nominal constructions. Moreover, the modification of deverbal NVEs, influenced by their internal structure, has implications for the syntax-semantics interface.

Let us wrap up the conclusions by considering how the findings answer the initial research questions this dissertation raised at the outset.

RQ1. Are there any divergences in the selection of the NVE by the LV across languages?

Some divergences have been found in terms of patterns related to the selection of the NVE, but these differences are strictly related to the general tendencies of the languages in the nominal domain.

RQ1a. If so, what are the differences in terms of determination and modification?

In this dissertation, the differences have been analysed especially regarding two aspects: determination and modification. On the one hand, regarding determination, all languages have NVEs predominantly introduced by a determiner; only Ger. *nehmen*-LVCs and Cat. *prendre*-LVCs ('take') select more bare nouns as NVEs than determined NVEs. This pattern is especially relevant in the case of German: the LV *nehmen* 'take' shows a different behaviour than the other two LVs (*geben* 'give' and *machen* 'make') and, thus, singles this verb out from the general tendency of the language and across languages.

On the other hand, regarding the modification of the NVE, English LVCs have a strong prevalence of modified NVEs which contrasts with the other three languages: Catalan and Spanish have around a third of all instances with modified NVEs, while German presents a lower presence. In line with previous studies, adjectival modification is the most common subclass of modification in all languages for *give* and *take,* whereas the LV *make* shows a strong presence of PP modification in the two Romance languages (Cat. *fer*-LVCs and Sp. *hacer*-LVCs). The particularities in the selection of PP modification in these languages can be traced to the transitive verbal base of the eventive noun, as in the case of Cat. *fer fotografia* 'make picture' or Sp. *hacer uso* 'make use'.

RQ1b. Does the absence of determination influence the structures within the NVE?

As shown in Chapter 6, only NVE introduced by a determiner can be passivized, pronominalized, interrogated and accept relative clauses with scope on the nominal element in the four languages. In contrast, bare nominals show more restrictions, as they do not accept passivization, although they are possible with *se*-passives in Catalan and Spanish – which are not passives morphologically; they do not undergo pronominalization except for the partitive *en* in Catalan; they cannot be interrogated; and relative clauses are only accepted as modifiers of the whole predicate. The behaviour of bare NVEs regarding topicalization differs across languages and language families. On the one hand, English does not accept it, while German allows it for both determined and bare nominals. On the other hand, Romance languages accept it with determined NVEs and only through the partitive *en* in Catalan when the topic is a bare noun.

The absence of determination has an influence on the argumental properties of the NVE: bare noun phrases are unable to function as canonical arguments. This phenomenon is, however, not restricted to LVCs, and is comparable to other constructions previously studied as pseudo-incorporated nominals (e.g. *tener piso* 'have flat' in Espinal & McNally, 2011, among others). Unlike the predicates analysed as pseudo-incorporated nominals, NVEs in LVCs show more flexibility in terms of modification. The only restriction found in this crosslinguistic study is limited to non-deverbal bare NVEs, which do not show restrictions in the kind of modification acceptable in English and German LVCs, but they do not accept qualitative modification in Catalan and Spanish LVCs, as already seen in (5) and (6).

Regarding the relationship between LVCs and their synthetic verbal counterpart (e.g. *take a shower – to shower*), the findings concentrate solely on Catalan data. In this case, the absence of the determiner facilitates the LVC to be interchangeable with a synthetic verb. More specifically, bare and unmodified NVEs are more often replaceable than determined NVEs with modification.

RQ1c. Is the degree of frequency of the LVC a further determining factor?

The degree of frequency of the LVC is not a clear determining factor, because it shows varying effects on modification. The corpus data has shown divergent findings in relation to the specific LV analysed: high-frequency LVCs are more modified in *give*-LVCs for all languages, while *make*-LVCs in low frequency lemmas. Moreover, *take*-LVCs differ across language families: the two Germanic languages have more modification in high and medium frequency, and the two Romance languages in low frequency LVCs.

The corpus data have shown that the most frequent LVCs do not present signs of lexicalization. Thus, their extended use in the language alongside their synthetic counterparts cannot be traced back to their co-occurring tendencies with certain modification subclasses, such as specific adjectives or the selection of PP modification.

RQ2. To which extent LVCs and their corresponding synthetic counterpart convey the same meaning?

The coexistence of LVCs with their synthetic counterparts has been explored only on the basis of Catalan LVCs. The corpus data has shown that Catalan LVCs are interchangeable by a synthetic verb in 57.89% of the instances, which is the majority of examples but there is a 42.11% of cases which cannot be replaced. In fact, these findings have shown that LVCs do provide a valid alternative to their corresponding synthetic verb and are, thus, not against the principle of economy in language (proposed by Zipf, 1949, and Martinet, 1955).

RQ2a. If there are differences, what are the syntactic, semantic and discourse differences between the LVCs and the corresponding synthetic verb?

In this case, the interchangeability of Catalan LVCs by their synthetic verbal counterparts can be explained at different linguistic levels: discourse, morphosyntax and semantics.

First, discourse plays a role in influencing interchangeability, particularly when elements of the construction belong to different clauses and there is cross-clausal reference.

Second, morphosyntactic properties, such as subject characteristics and the weight of the nominal element (determination, number, and modification), also influence interchangeability. More specifically, the corpus data has confirmed the findings of previous corpus-based studies on different languages: LVCs with bare nouns have proven to be statistically more likely to be interchangeable. Likewise, the cases with unmodified NVEs are also more probably interchangeable than LVCs with modification. The weight on the nominal element, hence, has been found as an important variable when analysing the interchangeability of Catalan LVCs with their synthetic verbal counterpart.

Finally, it has been found that semantic correlation between LVCs and synthetic counterparts is generally present, although it is not always exact.

RQ3. How can LVCs be structurally accounted for within a neo-constructionist approach to argument structure?

Taking the corpus findings into consideration, the typology proposal suggests that all LVCs behave like regular verbo-nominal combinations, with differences in cohesion due to the nature of the NVE rather than the light verb itself. Following an extended project analysis within the nominal domain (Grimshaw, 2005), the compositional hierarchy within the DP illustrates the combinatorial possibilities of the nominal in LVCs: the more projection in the structure, the less integration of elements, and the more independence of the nominal element. An exception to this analysis has been found in a very restricted used of *dar vuelta* followed by a DO in Rioplatense Spanish, which has proven to need an incorporation analysis to allow the selection of a second object.

Finally, this dissertation has contributed to the description of LVCs in four different languages with a corpus linguistics methodology combined with a theoretical analysis of these constructions. The combination of both perspectives has provided the possibility to contrast a broad picture of the data with a more fine-grained analysis of the findings from a theoretical perspective.

Bibliography

Corpora

COCA = Davies, Mark (2008). *The Corpus of Contemporary American English.* Available online at: https://www.english-corpora.org/coca/

CORPES XXI = Real Academia Española. 2013. *Corpus del Español del siglo XXI.* Available online at: https://www.rae.es/banco-de-datos/corpes-xxi

CTILC = Institut d'Estudis Catalans. 2005. *Corpus textual informatitzat de la llengua catalana.* Available online at: https://ctilc.iec.cat/scripts/

DWDS = Berlin-Brandenburgischen Akademie der Wissenschaften. 2006. *Digitales Wörterbuch der deutschen Sprache.* Available online at: https://www.dwds.de/

References

Abeillé, A. (1988). Parsing French with Tree Adjoining Grammar: Some linguistic accounts. In D. Vargha (Ed.), *Coling Budapest: Proceedings of the 12th International Conference on Computational Linguistics* (pp. 7-12), Association for Computational Linguistics. https://aclanthology.org/C88-1002.pdf

Abney, S. P. (1987). *The English noun phrase in its sentential aspect* [Doctoral dissertation]. Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/14638

Acedo Matellán, V. (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

Acedo-Matellán, V. (2014). Can we dance without doing a dance? Two opposite views on the integration of roots in the syntactic structure of the vP. In A. Bachrach, I. Roy, & C. Stockall (Eds.), *Structuring the Argument. Multidisciplinary research on verb argument structure* (pp. 23-43), John Benjamins. https://doi.org/10.1075/lfab.10.02ace

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

Acedo-Matellán, V. & Pineda, A. (2019). Light verb constructions in Basque and Romance. In A. Berro, B. Fernández & J. Ortiz de Urbina (Eds.), *Aligning grammars: Basque and Romance* (pp. 176-220), Brill. https://doi.org/10.1163/9789004395398_007

Alba-Salas, J. (2002). *Light verb constructions in Romance: A syntactic analysis* [Doctoral dissertation]. Cornell University.

Alba-Salas, J. (2004). Fare light verb constructions and Italian causatives: Understanding the differences. *Rivista di Linguistica*, *16*(2), 283-323.

Alexiadou, A. (2001). *Functional structure in nominals: Nominalization and ergativity* (Vol. 42). John Benjamins Publishing. https://doi.org/10.1075/la.42

Alexiadou, A., Iordăchioaia, G. & Soare, E. (2010). Number/aspect interactions in the syntax of nominalizations. *Journal of Linguistics, 46*(3). 537–574. http://doi.org/10.1017/S0022226710000058

Alexiadou, A. (2020, September 30). *Nominalization* [Invited Talk]. Abralin Ao Vivo. https://aovivo.abralin.org/en/lives/argument-structure-2/

Alexiadou, A., & Borer, H. (Eds.). (2020). *Nominalization: 50 Years on from Chomsky's Remarks* (Vol. 76). Oxford University Press. https://doi.org/10.1093/oso/9780198865544.001.0001

Alexiadou, A., & Grimshaw, J. (2008). Verbs, nouns and affixation. In F. Schäfer (Ed.), *Working Papers of the SFB 732 Incremental Specification in Context,* 01, 1-16.

Alonso Ramos, M. (2004). *Las construcciones con verbos de apoyo.* Visor Libros.

Alsina, A. (1992). On the Argument Structure of Causatives. *Linguistic Inquiry, 23*(4), 517–555. http://www.jstor.org/stable/4178789

Alsina, A. (1996). *The Role of Argument Structure in Grammar: Evidence from Romance*. CSLI Publications.

Alvarez-Morera, G. (2023). L'ús dels verbs de suport fer i donar en el català actual. In J. Marqués Messeguer & E. Aguilar Miró (Eds.), *(De) bat a bat: obertures i cruïlles en estudis recents sobre literatura i llengua catalanes.* Presses universitaires de Perpignan. https://doi.org/10.4000/books.pupvd.41594

Alvarez-Morera, G. (to appear). Light Verb Constructions at two ends of a continuum. In M. Bargalló Escrivà (Ed.), *Recerca en Humanitats 2022*. Publicacions Universitat Rovira i Virgili.

Arús Hita, J., & Martínez Caro, E. (2020). La construcción con verbo de apoyo *dar* en español y su comparación con la correspondiente *give* en inglés: implicaciones para la traducción. In J. Lavid Lopez (Ed.), *Contrastes inglés-español y traducción: estudios basados en corpus* (pp. 131-156). Guillermo Escolar Editor.

Baker, M. (1988). Theta theory and the syntax of applicatives in Chichewa. *Natural Language & Linguistic Theory,* 6, 353-389.

Baker, M. C. (2014). Pseudo Noun Incorporation as Covert Noun Incorporation: Linearization and Crosslinguistic Variation. *Language and Linguistics*, 15(1):5-46. https://doi.org/10.1177/1606822X13506154

Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., & Schneider, N. (2013). *Abstract Meaning Representation (AMR) 1.0 Specification*. http://www.isi.edu/~ulf/amr/help/amr-guidelines.pdf.

Barrie, M., & Mathieu, E. (2016). Noun incorporation and phrasal movement. *Natural Language & Linguistic Theory, 34*(1), 1-51. https://www.jstor.org/stable/43698466

Barrie, M. J. M., Li, A., Wiltschko, M., & Park, J. U. (2021). In Defense of DP (or KP). *Linguistic Research, 38*(2), 207-238. https://doi.org/10.17250/khisli.38.2.202106.002

Berenjian, S. R., & Rahimian, J. (2021). An Investigation into Lexical Aspect of Persian Light Verbs: zadan, dădan and xordan based on Huddleston and Pullum (2002). *International Journal of Linguistics, Literature and Translatio*n*, 4*(3), 01-16. https://doi.org/10.32996/ijllt.2021.4.3.1

Berlage, E. (2012). At the interface of grammaticalisation and lexicalisation: the case of take prisoner. *English Language & Linguistics, 16*(1), 35-55. https://doi.org/10.1017/S136067431100027X

Bernstein, J. B. (2001). The DP hypothesis: Identifying clausal properties in the nominal domain. In M. Baltin & C. Colins (Eds.), *The handbook of contemporary syntactic theory* (pp. 536-561). Blackwell Publishers. https://doi.org/10.1002/9780470756416.ch17

Biber, D. (1993). Representativeness in corpus design. *Literary and linguistic computing*, *8*(4), 243-257. https://doi.org/10.1093/llc/8.4.243

Biber, D., & Conrad, S. (2019). *Register, genre, and style.* Cambridge University Press. https://doi.org/10.1017/CBO9780511814358

Bilous, R. (2011). *Transitivité et marquage d'objet différentiel* [Doctoral dissertation]. University of Toronto. https://tspace.library.utoronto.ca/bitstream/1807/31692/1/Bilous_Rostyslav_201111_PhD_thesis.pdf

Blümel, A., & Holler, A. (2022). DP, NP, or neither? Contours of an unresolved debate. *Glossa: a journal of general linguistic*s, 7(1). https://doi.org/10.16995/glossa.8326

Bofarull, J. de; & Montserrat, A. (2001). *Praxi lingüística III. Documents administratius.* Universitat Rovira i Virgili, 20-21. https://www.llengues.urv.cat/media/upload/domain_81/arxius/praxi3.pdf

Bonet, S., & Solà, J. (1986). *Sintaxi generativa catalana.* Enciclopèdia Catalana.

Bonial, C. N. (2014). *Take a look at this! Form, function and productivity of English light verb constructions* [Doctoral dissertation]. University of Colorado at Boulder.

Bonial, C. & Pollard, K. A. (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

Borer, H. (2005). *Structuring Sense.* Oxford University Press.

Borer, H. (2013). *Structuring sense: Taking form* (vol. 3). Oxford University Press.

Borik, O., & Gehrke, B. (2015). An Introduction to the Syntax and Semantics of Pseudo-Incorporation. In O. Borik & B. Gehrke (Eds.), *The Syntax and Semantics of Pseudo-Incorporation*. Brill. https://doi.org/10.1163/9789004291089_002

Borthen, K. (2003). *Norwegian bare singulars* [Doctoral dissertation], Trondheim Norwegian University of Science and Technology. https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/242756

Bosque, I. (1999). El nombre común. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (vol. 1, pp. 3-75). Espasa Calpe.

Bosque, I. (2004). Combinatoria y significación. Algunas reflexiones. In I. Bosque (Dir.), *REDES. Diccionario combinatorio del español contemporáneo* (pp. lxxvii-clxxi). SM.

Bosque, I. (2012). Sobre el adjetivo mismo en las construcciones de dependencia interna. In T. Jiménez Juliá, B. López Meirama, V. Vázquez Rozas & A. Veiga (Eds.), *Cum corde et in nova grammatica. Estudios ofrecidos a Guillermo Rojo* (pp. 93-108). Servizo de Publicacións e Intercambio Científico da Universidade de Santiago de Compostela. http://hdl.handle.net/10347/12188

Bosque, I., & Picallo, C. (1996). Postnominal adjectives in Spanish DPs. *Journal of linguistics, 32*(2), 349-385. https://doi.org/10.1017/S0022226700015929

Bowern, C. (2008). The diachrony of complex predicates. *Diachronica, 25*(2), 161-185. https://doi.org/10.1075/dia.25.2.03bow

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

Brezina, V. (2018). *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899

Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change.* Cambridge University Press. https://doi.org/10.1017/CBO9780511615962

Brinton, L. J. (2008). Where lexis and grammar meet: Composite predicates in English. In M. J. López-Couso & E. Seoane (Eds.), *Theoretical and empirical issues in grammaticalization* (pp. 33-54). John Benjamins. https://doi.org/10.1075/tsl.77.04bri

Brito, A. M. (2010). Do European Portuguese and Spanish have the Double Object Construction? In *V Encuentro de Gramática Generativa*, Universidad Nacional del Comahue, Argentina.

Bruening, B. (2009). Selectional asymmetries between CP and DP suggest that the DP Hypothesis is wrong. In L. MacKenzie (Ed.), *U. Penn working papers in linguistics 15.1: Proceedings of the 32nd annual Penn linguistics colloquium*, 26–35.

Bruening, B. (2015). *Depictive secondary predicates, light verb give, and theories of double object constructions* [Manuscript]. University of Delaware.

Bruening, B. (2016). Light Verbs are Just Regular Verbs. *University of Pennsylvania Working Papers in Linguistics,* 22 (1), Article 7.

Bruening, B., Dinh, X., & Kim, L. (2018). Selection, idioms, and the structure of nominal phrases with and without classifiers. *Glossa: a journal of general linguistics,* 3(1), 42. https://doi.org/10.5334/gjgl.288

Bruening, B. (2020). The head of the nominal is N, not D: N-to-D movement, hybrid agreement, and conventionalized expressions. *Glossa: a journal of general linguistics*, 5(1). https://doi.org/10.5334/gjgl.1031

Bruening, B. (2022). Hebrew nominals do not require functional structure above the NP. *Glossa: a journal of general linguistics*, 7(1). https://doi.org/10.16995/glossa.5763

Brugman, C. (2001). Light verbs and polysemy. *Language sciences, 23*(4), 551-578. https://doi.org/10.1016/S0388-0001(00)00036-X

Bruker, A. (2011). *Entwurf und Realisierung von Lexikon-Einträgen für Funktionsverbgefüge der deutschen Sprache*. Diplomarbeiten Agentur.

Bustos Plaza, A. (2005). *Combinaciones verbonominales y lexicalización.* Peter Lang.

Butt, M. (1995). *The Structure of Complex Predicates in Urdu*. Stanford, California: CSLI Publications.

Butt, M. (2010). The light verb jungle: Still hacking away. In M. Amberber, B. Baker, M. Harvey (Eds.), *Complex Predicates: Cross-linguistic Perspectives on Event Structure* (pp.48-78). Cambridge University Press. https://doi.org/10.1017/CBO9780511712234.004

Butt, M., & Geuder, W. (2001). On the (semi) lexical status of light verbs. In N. Corver & H. van Riemsdijk (Eds.), *Semi-lexical categories: The function of content words and the content of function words* (pp. 323-370). Mouton de Gruyter. https://doi.org/10.1515/9783110874006.323

Butt, M., & Lahiri, A. (2013). Diachronic pertinacity of light verbs. *Lingua,* 135, 7– 29. https://doi.org/10.1016/j.lingua.2012.11.006

Camacho Coma, L. (2020). *Las Construciones con verbo de apoyo: el enriquecimiento del léxico a partir de la gramática de una lengua* [Master thesis]. Universitat de Girona. https://dugi-doc.udg.edu/handle/10256/18875

Carlson, G. (2003). No lack of determination. In L. Cheng & R. Sybesma (Eds.), *The second GLOT International state-of-the-article book: The latest in linguistics* (pp. 149-178). Mouton de Gruyter.

Castell, A. (2011). El tractament de les construccions amb verb suport en els diccionaris monolingües i bilingües de l'alemany i el català. *Zeitschrift für Katalanistik*, 24, 73-88.

Castillo Peña, C. (2013). *Las construcciones con verbos de apoyo en un corpus comparable de textos jurídicos*. EUT Edizioni Università di Trieste. http://hdl.handle.net/10077/9142

Cattell, R. (1984). 'Light' Verbs in English. In R. Cattell (Ed.), *Composite Predicates in English* (pp. 1-18). Brill.

Charnavel, I. (2019). On the sentence-internal reading of French le même ('the same'). In *Proceedings of* SALT, 20, 216-232. https://doi.org/10.3765/salt.v20i0.2567

Chen, W. T., Bonial, C., & Palmer, M. (2015). English light verb construction identification using lexical knowledge. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 29(1), 2375-2381. https://doi.org/10.1609/aaai.v29i1.9534

Cheng, W. (2012). *Exploring Corpus Linguistics: Language in Action*. Routledge.

Chierchia, G. (1998). Reference to kinds across language. *Natural language semantics, 6*(4), 339-405. https://www.jstor.org/stable/23748806

Choi, I., & Wechsler, S. (2002). Mixed Categories and Argument Transfer in the Korean light-verb Construction. In F. van Eynde, L. Hellan & D. Beermann (Eds.), *Proceedings of the 8th International HPSG Conference, Norwegian University of Science and Technology,* (pp. 103-120). CSLI Publications.

Choi, S. (2005). *Multiple Verb Constructions in Korean* [Doctoral dissertation], University of Sussex.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. The MIT Press.

Chomsky, N. (1986). *Barriers* (Vol. 13). The MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. The MIT Press.

Choupina, C., & Brito, A. M. (2018). Dar um beijo é a mesma coisa que dar um espirro?: para uma análise sintático-semântica de diferentes valores do verbo dar em português europeu. *A linguística em diálogo: volume comemorativo dos 40 anos do Centro de Linguística da Universidade do Porto,* 153-176. https://hdl.handle.net/10216/119712

Cinque, G. (1994). On the evidence for partial N-movement in the Romance DP. In G. Cinque (Ed.), *Paths towards Universal Grammar. Essays in honor of Richard S. Kayne.* Georgetown University Press, 85-110.

Cinque, G. (2005). Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry*, 36, 315– 323. https://www.jstor.org/stable/4179327

Cinque, G. (2010). *The syntax of adjectives: A comparative study* (Vol. 57). MIT press. https://doi.org/10.7551/mitpress/9780262014168.001.0001

Claridge, C. (2000). *Multi-Word verbs in Early Modern English: A corpus-based study.* Rodopi.

Coelho, L. C., & Sousa, V. V. (2017). A polissemia do verbo dar: uma análise funcionalista. *Entrepalavras*, 7(2), 74-92. http://dx.doi.org/10.22168/2237-6321.7.7.2.74-92

Colominas, C. (2001). *La representació semàntica de les construccions de suport des d'una perspectiva multilingual* [Doctoral dissertation], Universitat Autònoma de Barcelona. http://hdl.handle.net/10803/5252

Contreras, J. M., & Suñer, A. (2004). Los procesos de lexicalización. In E. Pérez Gaztelu, I. Zabala & L. Gràcia (Eds.), *Las fronteras de la composición en lenguas románicas y en vasco* (pp. 47-160). Universidad de Deusto.

Corver, N., & van Riemsdijk, H. (2001). Semi-Lexical Categories. In N. Corver & H. van Riemsdijk (Eds.), *Semi-Lexical Categories: The Function of Content Words and the Content of Function Words* (pp. 1-18). Mouton de Gruyter.

Cremades, E. (2014). *L'estructura argumental dels noms en català* [Doctoral dissertation]. Universitat Rovira i Virgili. http://hdl.handle.net/10803/306271

Cremades, E. (2016). Els arguments interns directes dels noms en català: una aproximació descriptiva. *Caplletra. Revista de Filologia*, 61, 91-123. https://raco.cat/index.php/Caplletra/article/view/313988

Cremades, E. (2017). *L'estructura argumental dels noms en català* (Vol. 465). Masarykova univerzita.

Cremades, E. (2018). Els noms eventuals en català. *Estudis Romànics*, 40, 117-144. https://raco.cat/index.php/Estudis/article/view/336564

Cuervo, M. C. (2003). *Datives at large* [Doctoral dissertation]. Massachusetts Institute of Technology. http://dspace.mit.edu/handle/1721.1/7991

Cuervo, M. C. (2010). La estructura de expresiones con verbos livianos y experimentante. *Cuadernos de la ALFAL*, 1, 194-206.

Cyrino, S. & Espinal, M. T. (2019). On the Syntax of Number in Romance. S*tudia Linguistica,* 74, 165-203. https://doi.org/10.1111/stul.12123

Davies, M. (2008). *The Corpus of Contemporary American English (COCA)*. Available online at https://www.english-corpora.org/coca/

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

De la Cruz, Y. (2020). *El verb tenir com a verb de suport: Anàlisi contrastiva a partir de dues novel·les catalanes contemporànies i la seva traducció al castellà* [Bachelor thesis], Universitat Rovira i Virgili.

De la Cruz, Y. (2021). *Anàlisi contrastiva de fer com a verb de suport en català i castellà* [Master thesis], Universitat Rovira i Virgili.

De Miguel, E. (2006). Tensión y equilibrio semántico entre nombres y verbos: el reparto de la tarea de predicar. In *Actas del XXXV Simposio Internacional de la Sociedad de Lingüıstica Espanola,* 1289-1313.

De Miguel, E. (2008). Construcciones con verbo de apoyo en español. De cómo entran los nombres en la órbita de los verbos. In I. Olza Moreno, M. Casado Velarde, R. González Ruiz (Eds.), *Actas del XXXVII Simposio Internacional de la Sociedad Española de Lingüística*. Servicio de Publicaciones de la Universidad de Navarra, 567-578.

De Miguel, E. (2011). En qué consiste ser verbo de apoyo. In M. V. Escandell Vidal, M. Leonetti & C. Sánchez López (Ed.), *60 problemas de gramática dedicados a Ignacio Bosque*. Madrid: Akal, 139-147.

deMena Travis, L. (2014). The integration, proliferation, and expansion of functional categories. In A. Carnie, D. Siddiqi & Y. Sato (Eds.), *The Routledge Handbook of Syntax Routledge* (pp. 42-64). Routledge.

Demonte, V. (1995). Dative alternation in Spanish. *Probus,* 7, 5-30.

DIEC = Institut d'Estudis Catalans. *Diccionari de la llengua catalana de l'IEC*. Institut d'Estudis Catalans. https://dlc.iec.cat/

Di Tullio, Á., Saab, A., & Zdrojewski, P. (2019). Clitic Doubling in a Doubling World: The Case of Argentinean Spanish Reconsidered. In Á. Gallego (Ed.), *The Syntactic Variation of Spanish Dialects*, Oxford University Press, 215-244. https://doi.org/10.1093/oso/9780190634797.001.0001

Dobrovie-Sorin, C., Bleam, T., & Espinal, M. T. (2006). Bare nouns, number and types of incorporation. In S. Vogeleer & L. Tasmowski (Ed.), *Non-definiteness and plurality*, John Benjamins, 51-79.

DRAE = REAL ACADEMIA ESPAÑOLA. *Diccionario de la lengua española* (23th ed.), [version 23.6]. https://dle.rae.es

DUDEN = Dudenredaktion. (2020). *Duden online*. Cornelsen Verlag. https://www.duden.de/

Durrant, P. & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations?, *International Review of Applied Linguistics*, 47, 157–177. https://doi.org/10.1515/iral.2009.007

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and evaluating language corpora: A practical framework for corpus representativeness.* Cambridge University Press. https://doi.org/10.1017/9781316584880

Egbert, J., Larsson, T., & Biber, D. (2020). *Doing linguistics with a corpus: Methodological considerations for the everyday user.* Cambridge University Press. https://doi.org/10.1017/9781108888790

Eisenberg, P., Peters, J., Gallmann, P., Fabricius-Hansen, C., Nübling, D., Barz, I., Fritz, T. A. & Fiehler, R. (2005). *Duden: Die Grammatik* (7th edition), vol. 4. Dudenverlag

Elenbaas, M. (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. T. (2013). Bare nominals, bare predicates. In A. Wall & J. Kabatek (Eds.), *New perspectives on bare noun phrases in romance and beyond* (pp. 63-94). John Benjamins. https://doi.org/10.1075/slcs.141

Espinal, M. T. & McNally, L. (2011). Bare nominals and incorporating verbs in Catalan and Spanish. *Journal of Linguistics,* 47, 87-128. https://doi.org/10.1017/S0022226710000228

Espinal, M. T., & Dobrovie-Sorin, C. (2006). Tipología semántica de los nombres escuetos. El caso particular de los nombres escuetos singulares contables. *Andolin Gogoan: Essays in Honour of Prof. Eguzkitza* (pp. 269-285). Universidad del País Vasco.

Fábregas, A. (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

Fábregas, A. (2021). SE in Spanish: properties, structures and analyses. *Borealis. An International Journal of Hispanic Linguistics*. *10*(2), 1-236. https://doi.org/10.7557/1.10.2.5934

Farkas, D., & de Swart, H. (2003). *The semantics of incorporation: From argument structure to discourse transparency.* CSLI Publications.

Fleischhauer, J. (2020). Predicative multi-word expressions in Persian. *Proceedings of the 34th Pacific Asia Conference on Language, Information and Computation*, 552–561. https://aclanthology.org/2020.paclic-1.63

Fleischhauer, J. (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. https://doi.org/10.18653/v1/2021.mwe-1.8

Fleischhauer, J. & Gamerschlag, T. (2019). Deriving the meaning of light verb constructions - a frame account of German stehen 'stand'. In C. Juchem-Grundmann, M. Pleyer & M. Pleyer (Eds.), *Yearbook of the German Cognitive Linguistics Association* (pp. 137-156). Mouton de Gruyter. https://doi.org/10.1515/gcla-2019-0009

Fleischhauer, J., & Neisani, M. (2020). Adverbial and attributive modification of Persian separable light verb constructions. *Journal of Linguistics,* 56(1), 45-85. https://doi.org/10.1017/S0022226718000646

Folli, R., Harley, H., & Karimi, S. (2005). Determinants of event type in Persian complex predicates. *Lingua*, 115(10), 1365-1401.

Folli, R., & Harley, H. (2013). The syntax of argument structure: Evidence from Italian complex predicates. *Journal of Linguistics,* *49*(1), 93-125. https://doi.org/10.1017/S0022226712000072

Fullana, C.; Alvarez-Morera, G. & Oltra-Massuet, I. (2022). Unergative Cognate Structures across English and Romance. In F. Gallardo del Puerto, M-C. Camus Camus & J.A. González-López (Eds.), *Moving beyond the Pandemic: English and American Studies in Spain* (pp. 65-77). Ediciones Universidad de Cantabria. https://doi.org/10.22429/Euc2022.034

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

García Pérez, R. (2007). *¿Qué hacíamos y qué hacemos?: el verbo hacer en la historia del español,* Cilengua.

Garcia-Vega, M., & Machonis. P. (2014). The Support Verb take. In F. Kakoyanni-Doa (Ed.), *Penser le Lexique-Grammaire. Perspectives actuelles* (p. 111-123). Honoré Champion.

Gianollo, C., von Heusinger, K., & Napoli, M. (2021). Chapter 1 Reference and Quantification in Nominal Phrases: The Current Landscape and the Way Ahead. In C. Gianollo, K. von Heusinger & M. Napoli (Eds.), *Determiners and Quantifiers*. Brill. https://doi.org/10.1163/9789004473324_002

GIEC = Institut d'Estudis Catalans (2016).*Gramàtica de la llengua catalana.* Institut d'Estudis Catalans.

Ginebra, J. (2003). Fraseologia, concurrències lèxiques i llengua estandard. In M. À. Pradilla (Ed.), *Identitat lingüística i estandardització* (pp. 7-55). Cossetània.

Ginebra, J. (2006). Sobre la riquesa lèxica de Verdaguer: una primera prospecció quantitativa. *Anuari Verdaguer,* 14, 311-324. https://raco.cat/index.php/AnuariVerdaguer/article/view/68220

Ginebra, J. (2008). Els verbs de suport en català i en anglès: estudi contrastiu a partir d'un petit corpus paral·lel. *Els Marges: revista de llengua i literatura*, 85, 53-73. https://raco.cat/index.php/Marges/article/view/142237

Ginebra, J. (2017). *La nova normativa de l'Institut d'Estudis Catalans: guia pràctica*, Publicacions Universitat Rovira i Virgili.

Ginebra, J. & Navarro, P. (2015). Concurrències lèxiques en català i en espanyol: uns quants contrastos. In À. Martín Escribà, A. Piquer Vidal & F. 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, pp. 217-228). Publicacions de l'Abadia de Montserrat.

Giparaitė, J. (2016). Complementation of light verb constructions in World Englishness: A corpus-based study. *Žmogus ir žodis,* 18(3), 19-39.

Giry-Schneider, J. (1987). *Les prédicats nominaux en français: les phrases simples à verbe support (Vol. 18)*. Librairie Droz.

Gonçalves, A., Cunha, L. F., Miguel, M., Silvano, M. D. P., & Silva, M. D. F. H. D. (2010). Propriedades predicativas dos verbos leves: estrutura argumental e eventiva. *XXV Encontro Nacional da Associação Portuguesa de Linguística: Textos seleccionados*, 449-464.

Gonçalves, R. (2016). *Construções ditransitivas no Português de São Tomé* [Doctoral dissertation]. Universidade de Lisboa. http://hdl.handle.net/10451/26409

Grimshaw, J. (1990). *Argument structure*. MIT Press.

Grimshaw, J. (1991). *Extended projection* [Unpublished paper]. Brandeis University.

Grimshaw, J. (2005). *Words and Structure*. Center for the Study of Language and Information.

Grimshaw, J. & Mester. A. (1988). Light Verbs and Ө-Marking. *Linguistic Inquiry, 19*(2), 205-232.

Gross, M. (1989). Les expressions figées en Français. *Technical Report*, University Paris 7.

Hale, K. & Keyser, S. J. (1993). On argument structure and the lexical expression of syntactic relations. In K. Hale & S. J. Keyser (Eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (pp. 53–109). MIT Press.

Hale, K. & Keyser, S. J. (2002). *Prolegomenon to a theory of argument structure.* MIT Press.

Halle, M., & Marantz, A. (1993). Distributed Morphology and the Pieces of Inflection. In K. Hale & S. J. Keyser (eds.). *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (pp. 111-176). MIT Press.

Harley, H. (2002). Possession and the double object construction. *Yearbook of Linguistic Variation*, 2, 29-68.

Harley, H. (2005). How do verbs get their names? The aktionsart of denominal verbs and the ontology of verb roots in English.' In N. Erteschik-Shir & T. Rapaport (Eds.), *The Syntax of Aspect* (pp. 42-64). Oxford University Press.

Heine, B. (1993). *Auxiliaries: Cognitive forces and grammaticalizatio*n. Oxford University Press.

Helbig, G. (1984). Probleme der Beschreibung von Funktionsverbgefügen im Deutschen. In *Studien zur deutschen Syntax* (pp. 163–188). VEB Verlag Enzyklopädie.

Höhle, T. N. (1982). Explikationen für „normale Betonung" und „normale Wortstellung". In Werner Abraham (ed.). *Satzglieder im Deutschen – Vorschläge zur syntaktischen, semantischen und pragmatischen Fundierung (Studien zur deutschen Grammatik 15,* pp. 75–153). Gunter Narr Verlag.

Hopper, P. J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge University Press.

Huddleston, R. & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.

Hundt, M., Nesselhauf, N., & Biewer, C. (2007). *Corpus linguistics and the Web*. Rodopi.

Husić, H. (2020). A vagueness based analysis of abstract nouns. *Proceedings of Sinn Und Bedeutung*, *24*(1), 359–376. https://doi.org/10.18148/sub/2020.v24i1.872

Iordăchioaia, G. (2020). D and N are different nominalizers. *Glossa: a journal of general linguistics*, *5*(1). https://doi.org/10.5334/gjgl.1111

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. MIT. Press.

Jardón Pérez, N. (2022). *Pluractional Perfects: Anatomy of a Construction in Eonavian Spanish* [Doctoral dissertation]. UiT The Arctic University of Norway. https://hdl.handle.net/10037/24059

Jespersen, O. (1954). *A Modern English Grammar* (vol. 4). Barnes and Noble.

Jiménez Martínez, M. I. (2016). *Colocaciones y verbos soporte en latín: semántica y sintaxis del verbo pono* [Doctoral dissertation]. Universidad Complutense de Madrid. https://hdl.handle.net/20.500.14352/22492

Kambanaros, M., Grohman, K. K., Michaelides, M., & Theodorou, E. (2014). On the nature of verb–noun dissociations in bilectal SLI: A psycholinguistic perspective from Greek. *Bilingualism: Language and Cognition*, *17*(1), 169-188. https://doi.org/10.1017/S1366728913000035

Kearns, K. (1988). *Light Verbs in English* [Doctoral dissertation]. Massachusetts Institute of Technology.

König, E. (2012). Contrastive linguistics and language comparison. *Languages in Contrast, 12*(1), 3-26. https://doi.org/10.1075/lic.12.1.02kon

Kornfeld, L. M. (2004). *Formación de palabras en la sintaxis desde la perspectiva de la Morfología Distribuida* [Doctoral dissertation]. Universidad de Buenos Aires. http://repositorio.filo.uba.ar/handle/filodigital/1567

Kratzer, A. (1996). Severing the external argument from its verb. In J. Rooryck & L. Zaring (Eds.), *Phrase Structure and the Lexicon. Studies in Natural Language and Linguistic Theory* (vol. 33, pp. 109-137). Springer.

Krifka, M. (1998). Scope Inversion under the Rise-Fall Contour in German. *Linguistic Inquiry*, 29 (1), 75-112. https://doi.org/10.1162/002438998553662

Laca, B. (1999). Presencia y ausencia de determinante. In I. Bosque & V. Demonte (Dir.), *Gramática descriptiva de la lengua española* (pp. 891-928). Espasa Calpe.

Landau, I. (2008). Two routes of control: Evidence from case transmission in Russian. *Natural Language & Linguistic Theory*, 26, 877-924. https://doi.org/10.1007/s11049-008-9054-0

Landau, I. (2013). *Control in generative grammar: A research companion.* Cambridge University Press. https://doi.org/10.1017/CBO9781139061858

Landau, I. (2015). *A two-tiered theory of control* (Vol. 71). MIT Press. https://www.jstor.org/stable/j.ctt17kk9r9

Laufer, B., & Waldman, T. (2011). Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning,* 61, 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Le Bruyn, B. S. W., de Swart, H. E., Zwarts, J., & Aronoff, M. (2017). Bare nominals. In *Linguistics Oxford Research Encyclopedias.* https://doi.org/10.1093/acrefore/9780199384655.013.399

Leech, G. (1991). The state of the art in corpus linguistics. In K. Aijmer and B. Altenberg (eds.), *English Corpus Linguistics* (pp. 8-29). Longman.

Leech, G., Hundt, M., Mair, C., & Smith, N. (2009). *Change in Contemporary English. A Grammatical Study*. Cambridge University Press. https://doi.org/10.1017/CBO9780511642210

Levin, B., & Rappaport Hovav, M. (2005). *Argument realization* (Vol. 10). Cambridge university press. https://doi.org/10.1017/CBO9780511610479

Levin, M., & Ström Herold, J. (2015). Give and Take: A contrastive study of light verb constructions in English, German and Swedish. In S. Oksefjell Ebeling and H. Hasselgård (eds.), *Cross-Linguistic Perspectives on Verb Constructions* (pp. 144-168). Cambridge Scholars Publishing.

Levshina, N. (2015). *How to Do Linguistics with R: Data exploration and statistical analysis*. John Benjamins. https://doi.org/10.1075/z.195

L'Homme, M. C. (2007). Using explanatory and combinatorial lexicology to describe terms. In L. Wanner (Ed.), *Selected Lexical and Grammatical Topics in the Meaning-Text Theory. In Honour of Igor Mel'čuk* (pp. 167-202).

Li, J., & Schmitt, N. (2010). The development of collocation use in academic texts by advanced L2 learners: A multiple case study approach. In Wood, D. (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 2–46). Continuum.

Lin, T.-H. (2001). *Light Verb Syntax and the Theory of Phrase Structure* [Doctoral dissertation]. University of California, Irvine.

Live, A. H. (1973). The take-have phrasal in English. *Linguistics,* 95, 31–50.

Longobardi, G. (1994). Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form, *Linguist Inquiry, 25*(4), 609-665.

Longobardi, G. (2001). How comparative is semantics? A unified parametric theory of bare nouns and proper names. *Natural language semantics*, *9*(4), 335-369.

Longobardi, G. (2005). Toward a Unified Grammar of Reference. *Zeitschrift für Sprachwissenschaft*, 24, 5-44.

Lopes, A. C. M. (2016). A polifuncionalidade de "mesmo" no português europeu contemporâneo, *Diacrítica*, *32*(1), 43-67. https://doi.org/10.21814/diacritica.93

López, C. S. (2002). Las construcciones con se. Estado de la cuestión. Madrid: Visor, 13-163.

Macfarland, T. (1995). *Cognate Objects and the Argument/Adjunct Distinction in English* [Doctoral dissertation]. Illinois: Northwestern University.

Massam, D. (2001). Pseudo Noun Incorporation in Niuean. *Natural Language & Linguistic Theory*, 19, 153–197.

Massam, D. (1990). Cognate Objects as Thematic Objects. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 35*(2), 161–90.

Marantz, A. (1984). *On the nature of grammatical relations*. MIT Press.

Marantz, A. (1991). Case and licensing. In G. Westphal, B. Ao & H.-R. Chae (Eds.). *Proceedings of ESCOL 1991*, 234–253.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics*, 4(2), 201-225.

Marantz, A. (2001). *Words and things*. MIT Press.

Marantz, A. (2011). Syntactic approaches to argument structure without incorporation: Doing the (anti-lexicalist) dance without doing the dance. Presented at *Structuring the Argument: A multidisciplinary workshop on the mental represenation of verbal argument structure. Structures Formelles du Langage.* Université Paris 8/CNRS.

Martínez, C. M. (2012). Aspectos de variación de las locuciones verbales con sustantivos escuetos en la historia del español. In *Actas del VIII Congreso Internacional de Historia de la Lengua Española: Santiago de Compostela*, 14-18 de septiembre de 2009 (pp. 959-970). Meubook.

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

Martínez Caro, E., & Arús Hita, J. (2020). Give as a light verb. *Functions of Language, 27*(3), 280-306. https://doi.org/10.1075/fol.16036.mar

Martinet, A. (1955). *Economie des changements phonétiques traité de phonologie diachronique.* A. Francke.

Masullo, P. J. (1992). *Incorporation and case theory in Spanish: a cross-linguistic perspective* [Doctoral dissertation]. University of Washington.

Masullo, P. J. (1996). Los sintagmas nominales sin determinante: una propuesta incorporacionista. In I. Bosque (Ed.), *El sustantivo sin determinación: la ausencia del determinante en la lengua espanola* (pp. 169-200). Visor.

Mateu, J. (2002). *Argument structure: Relational construal at the syntax-semantics interface* [Doctoral dissertation]. Universitat Autònoma de Barcelona.

Mateu, J., & Acedo-Matellán, V. (2012). 9: The Manner/Result Complementarity Revisited: A Syntactic Approach. In M. C. Cuervo & Y. Roberge (Eds.), *The end of argument structure* (pp. 209-228). Brill. https://doi.org/10.1163/9781780523774

Matsumoto, Y. (1996). A syntactic account of light verb phenomena in Japanese. *Journal of East Asian Linguistics*, 5(2), 107-149.

Matute Martínez, C. (2012). Aspectos de variación de las locuciones verbales con sustantivos escuetos en la historia del español. In *Actas del VIII Congreso Internacional de Historia de la Lengua Española* (pp. 959-970). Meubook.

McEnery, T., & Hardie, A. (2012) Corpus Linguistics: Method, Theory and Practice. *Corpus Linguistics and Linguistic Theory*, 9 (2), 293-303.

McEnery, T., & Wilson A. (1996/2001). *Corpus linguistics.* Edinburgh University Press.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Taylor & Francis.

McGinnis, M. (2008). Applicatives. *Language and Linguistics Compass*, 2(6), 1225-1245.

Mehl, S. (2018). Corpus onomasiology in world Englishes and the concrete verbs make and give. *World Englishes*, 37(2), 185-206. https://doi.org/10.1111/weng.12297

Mel'čuk, I. (1997). *Vers une linguistique Sens-Texte*. Collègue de France.

Mel'čuk, I. (1998). Collocations and lexical functions. In Cowie (Ed.), *Phraseology. Theory, analysis, and applications* (pp. 23-53). Oxford University Press.

Melloni, C. (2011). *Event and result nominals.* Peter Lang AG.

Mendívil-Giró, J. (1999). *Las palabras disgregadas: sintaxis de las expresiones idiomáticas y los predicados complejos.* Prensas Universitarias de Zaragoza.

Mendívil-Giró, J. (2019a). How Much Data Does Linguistic Theory Need? On the Tolerance Principle of Linguistic Theorizing. *Front. Commun.* 3, 62. https://doi.org/10.3389/fcomm.2018.00062

Mendívil-Giró, J. (2019b). If everything is syntax, why are words so important? An a-morphous but non-lexicalist approach. *Linguistics,* 57(5), 1161-1215. https://doi.org/10.1515/ling-2019-0025

Mihajlovska, K. (2017). *Enseñanza/aprendizaje de algunas construcciones basadas en los verbos soporte hacer, dar y tener en E/LE: elaboración de material didáctico destinado a aprendices macedonios* [Doctoral dissertation], Universitat Autònoma de Barcelona. http://hdl.handle.net/10803/456038

Mithun, M. (1984). The evolution of noun incorporation. *Language*, *60*(4), 847-894.

Mohanan, T. (2017). Grammatical and Light Verbs. *The Wiley Blackwell Companion to Syntax* (2ⁿᵈ Edition), 1-27. https://doi.org/10.1002/9781118358733.wbsyncom016

Montero Gálvez, S. (2014). Los nombres escuetos y el artículo Ø. *Verba hispánica*, *22*(1), 37-58.

Montolío Durán E. (2011). *Estudio de Campo: lenguaje escrito. Comisión para la modernización del lenguaje jurídico.* Ministerio de Justicia.

Montserrat, S. (2014). Dar/donar i fer + substantiu psicològic en català antic (ss. xiii-xvi): Una aproximació basada en la gramàtica de construccions. *Anuari de Filologia. Estudis de Lingüística*, 4, 157-183. https://doi.org/10.1344/AFEL2014.1.9

Müller, S. (2023). *Germanic syntax: A constraint-based view* (Textbooks in Language Sciences 12). Language Science Press. https://doi.org/10.5281/zenodo.7733033

Myler, N. (2016). *Building and interpreting possession sentences*. MIT press. https://doi.org/10.7551/mitpress/9780262034913.001.0001

Nagy, I., Rácz, A., & Vincze, V. (2020). Detecting light verb constructions across languages. *Natural Language Engineering*, *26*(3), 319-348. https://doi.org/10.1017/S1351324919000330

Nolan, B. (2014). Complex predicates and light verb constructions in Modern Irish. *Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics*, *27*(1), 140-167. https://doi.org/10.1075/resla.27.1.06nol

Norvig, P., & Lakoff, G. (1987). Taking: A study in lexical network theory. In *Annual Meeting of the Berkeley Linguistics Society,* Vol. 13, 195-206.

Nygård, G. (2021). *La relación semántica entre verbo ligero y verbo léxicamente pleno: Un estudio comparado sobre los verbos echar y soltar* [Doctoral dissertation], Norwegian University of Science and Technology. https://hdl.handle.net/11250/2758927

OED = *Oxford English Dictionary* (2016). Oxford University. https://www.oed.com/

Oggiani, C. (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, C. (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

Ortiz de Ciscomani, R. M. (2006). La bitransitividad. In C. Company Company (dir.), *Sintaxis histórica de la lengua española. Primera parte: La frase verbal*, vol. 1. Universidad Nacional Autónoma de México/Fondo de Cultura Económica, 575-668.

Oxford, W. (2010). *Same, different, and other: The microsyntax of identity adjectives* [Doctoral dissertation]. University of Toronto.

Oyharçabal, B.(2006). Basque Light Verb Constructions. *ASJU XL,* 787–806.

Pantcheva, M. (2009). First phase syntax of Persian complex predicates: Argument structure and telicity. *Journal of South Asian Linguistics*, *2*(1), 53–72.

Pereltsvaig, A. (1999). Cognate Objects in Russian: Is the Notion 'Cognate' Relevant for Syntax?. *Canadian Journal of Linguistics/Revue Canadienne de Linguistique, 44*(3), 267–91.

Pereltsvaig, A. (2002). Cognate Objects in Modern and Biblical Hebrew. In J. Ouhalla and U. Shlonsky (Eds.), *Themes in Arabic and Hebrew Syntax. Studies in Natural Language and Linguistic Theory,* (Vol 53, pp. 107-136). Springer.

Perea Irigoyen, D., & Sánchez Róo, S. (2013). Trayendo a la luz a los verbos ligeros del tipo 'hacer'en inglés y español. *CULCyT: Cultura Científica y Tecnológica*, *10*(49), 93-101.

Pérez Serrano, M. (2017). Colocaciones con verbos delexicalizados, ¿son más difíciles de aprender? *Language design: journal of theoretical and experimental linguistics*, 19, 127-148.

Picallo, M. C. (1991). Nominals and nominalizations in Catalan. *Probus, 3*(3), 279-316. https://doi.org/10.1515/prbs.1991.3.3.279

Picallo, M. C. (1994). Catalan possessive pronouns: the Avoid Pronoun Principle revisited. *Natural Language & Linguistic Theory*, *12*(2), 259-299.

Picallo, C. (2006). Some notes on grammatical gender and l-pronouns. In K. von Heusinger, G. A. Kaiser & E. Stark (Eds.), *Proceedings of the workshop 'Specificity and the evolution/emergence of nominal determination systems in Romance'* (pp. 107–121) Universität Konstanz.

Piera, C., & Varela, S. (1999). "Relaciones entre morfología y sintaxis". In V. Demonte & I. Bosque (Eds.), *Gramática descriptiva de la lengua española* (Vol. 3, pp. 4367-4422). Espasa Calpe.

Pineda, A. (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

Pineda, A. (2015). Del Datiu a l'acusatiu. Un canvi sintàctic en procés en llengües romàniques i basc. *Llengua i literatura*, 25, 73-98. https://raco.cat/index.php/LlenguaLiteratura/article/view/298801

Pineda, A. (2020). From dative to accusative. An ongoing syntactic change in Romance. *Probus*, 32(1), 129-173. https://doi.org/10.1515/probus-2019-0001

Plante, S. (2014). Types of nominalized verbs in light verb constructions. *McGill Working Papers in Linguistics*, *24*(1), 81-91.

Pollard, C., & Sag, I. (1987). *Information-Based Syntax and Semantic, Volume 1: Fundamentals* (CSLI Lecture Notes Series N. 13). CSLI Publications.

Porzig, W. (1957). *Das Wunder der Sprache.* Francke.

Pullum, G. K. (2017). Theory, data, and the epistemology of syntax. In M. Konopk & A. Wöllstein (Eds.), *Grammatische Variation: Empirische Zugänge und theorische Modellierung (Institut Für Deutsche Sprache, Jahrbuch 2016,* pp. 283–298). De Gruyter Mouton.

Pylkkänen, L. (2008). *Introducing arguments* (Vol. 49). MIT press.

Rácz, A., Nagy, I., & Vincze, V. (2014). 4FX: Light verb constructions in a multilingual parallel corpus. *Proceedings of the Ninth International Conference on Language Resources and Evaluation.* ELRA. http://www.lrec-conf.org/proceedings/lrec2014/pdf/331_Paper.pdf

RAE-ASALE (2009). *Nueva gramática de la lengua española.* Espasa Calpe.

Ramchand, G. (2008). *Verb meaning and the lexicon: A first-phase syntax* (Vol. 116). Cambridge University Press.

Ramchand, G. (2014). On structural meaning vs conceptual meaning in verb semantics. *Linguistic Analysis, 39*(1–2), 207–243.

Real Puigdollers, C. (2008). The Nature of Cognate Objects. A Syntactic Approach. *Proceedings ConSOLE XVI*, 157–78.

Real Puigdollers, C. (2013). *Lexicalization by Phase: The Role of Prepositions in Argument Structure and Its Cross-Linguistic Variation* [Doctoral dissertation]. Universitat Autònoma de Barcelona. http://hdl.handle.net/10803/120181

Reid, N. (2002). Sit Right Down the Back: Serialized Posture Verbs in Ngan'gityemerri and other Northern Australian Languages. In John Newman (ed.), *The Linguistics of Sitting, Standing and Lying* (pp. 293–268). John Benjamins.

Reinhart, T. (2006). *Interface strategies*. MIT Press.

Resnik, G. (2010). Derivación e interacción de rasgos: la delimitación en nombres y verbos derivados en español. In *Actas del V Encuentro de Gramática Generativa,* (pp. 405-421). Universidad Nacional del Comahue.

Richards, N. (2001). An Idiomatic Argument for Lexical Decomposition. *Linguistic Inquiry*, *32*(1), 183-192.

Rinaldi, M. G. (2018). *Bare Singulars and So-Called Bare Singulars* [Doctoral dissertation]. Queen Mary University of London. https://qmro.qmul.ac.uk/xmlui/handle/123456789/56405

Ritter, E. (1988). A head-movement approach to construct-state noun phrases. *Linguistics*, *26*(6), 909-930.

Ritter, E. (1991). Two Functional Categories in Noun Phrases: Evidence from Modern Hebrew. In *Syntax and Semantics*. Brill. https://doi.org/10.1163/9789004373198_004

Romera Martín, C. (2017). *Construcciones con verbo de apoyo. Caracterización en español y comparativa con el catalán* [Bachelor thesis]. Universitat Autònoma de Barcelona.

Romero, J. (1997). *Construcciones de doble objeto y gramática universal* [Doctoral dissertation]. Universidad Autónoma de Madrid

Ronan, P., & Schneider, G. (2015). Determining light verb constructions in contemporary British and Irish English. *International Journal of Corpus Linguistics*, *20*(3), 326-354. https://doi.org/10.1075/ijcl.20.3.03ron

Ronan, P. (2019). Simple versus Light Verb Constructions in Late Modern Irish English Correspondence: A Qualitative and Quantitative Analysis. *Studia Neophilologica*, *91*(1), 31-48. https://doi.org/10.1080/00393274.2019.1578182

Roodenburg, J. (2010). Plurality from a cross-linguistic perspective: the existence of plural argument supporting nominalizations in French. *Lingue e linguaggio*, *9*(1), 41-64.

Rosen, S. (1989). *Argument Structure and Complex Predicates* [Doctoral dissertation]. Brandeis University.

Rosselló, J. (2002). El SV, I: Verb I arguments verbals. In J. Solà, M. R. Lloret, J. Mascaró & M. Pérez Saldanya (Eds.), *Gramàtica del català contemporani* (vol. 3, pp. 1853-1949). Empúries.

Rothstein, S. (2010). Counting and the mass/count distinction. *Journal of semantics*, *27*(3), 343-397.

Rug, W., & Tomaszewski, A. (1993). *Grammatik mit Sinn und Verstand: 20 Kapitel deutsche Grammatik für Fortgeschrittene*. Klett.

Salzmann, M. (2020). The NP vs. DP debate. Why previous arguments are inconclusive and what a good argument could look like. Evidence from agreement with hybrid nouns. *Glossa: a journal of general linguistics*, *5*(1). https://doi.org/10.5334/gjgl.1123

Sánchez Rufat, A. (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

Sanromán Vilas, B. (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.

Sanromán Vilas, B. (2011). The unbearable lightness of light verbs. Are they semantically empty verbs. In *Proceedings of the 5th International Conference on Meaning-Text Theory,* 253-263.

Sanromán Vilas, B. (2013). Construcciones con verbo de apoyo cuasisinónimas. *Neuphilologische Mitteilungen*, *114*(4), 387-421. https://www.jstor.org/stable/43346343

Saromán Vilas, B. (2014). La alternancia dar/hacer en construcciones con verbo de apoyo y nombre de comunicación. *Borealis–An International Journal of Hispanic Linguistics*, *3*(2), 185-222. https://doi.org/10.7557/1.3.2.2946

Sanromán Vilas, B. (2015). The role of determiners in Spanish light verb constructions. *Word, 61*(2), 178-200. https://doi.org/10.1080/00437956.2015.1033176

Sanromán Vilas, B. (2017). ¿Es posible definir un verbo ligero?. *Lingua Americana*, 41, 17-46. http://hdl.handle.net/10138/240377

Schmid, B. (2016). Els usos del verb fer i la seva traducció a l'alemany. *Zeitschrift für Katalanistik,* 29, 115-134.

Schultze-Berndt, E. (2000). *Simple and Complex Predicates in Jaminjung* [Doctoral dissertation]. University of Nijmegen.

Seiss, M. (2009) On the difference between auxiliaries, serial verbs and light verbs. In M. Butt & T. H. King (Eds.), *Proceedings of the LFG09 Conference* (pp. 502–519). CSLI Publications.

Shen, L. (2004). Aspect agreement and light verbs in Chinese: A comparison with Japanese. *Journal of East Asian Linguistics*, *13*(2), 141-179.

Shomoossi, N., & Shomoossi, A. (2012). A review of light verb constructions in English and Persian complex predicates. *Acta Linguistica*, 6(2).

Slabakova, R., Kempchinsky, P., & Rothman, J. (2012). Clitic-doubled left dislocation and focus fronting in L2 Spanish: A case of successful acquisition at the syntax–discourse interface. *Second Language Research,* 28(3), 319-343. https://doi.org/10.1177/0267658312447612

Sleeman, P., & Brito, A. M. (2010b). Nominalization, event, aspect and argument structure. In M. Duguine, S. Huidobro, & N. Madariaga (Eds.), *Argument structure and syntactic relations A cross-linguistic perspective* (pp. 113-130). John Benjamins. https://doi.org/10.1075/la.158.07sle

Speelman, D. (2014). Logistic regression. A confirmatory technique for comparisons in corpus linguistics. In D. Glynn & J. A. Robinson (Eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy* (pp. 365-389). John Benjamins. https://doi.org/10.1075/hcp.43.18spe

Stone, M. S. (2016). *The difference between bucket-kicking and kicking the bucket: Understanding idiom flexibility* [Doctoral Dissertation]. University of Arizona at Tucson.

Storrer, A. (2007). Corpus-based investigations on German support verb constructions. In C. Fellbaum (Ed.), *Collocations and Idioms: Linguistic, lexicographic, and computational aspects.* Continuum Press.

Stvan, L. S. (2009). Semantic incorporation as an account for some bare singular count noun uses in English. *Lingua, 119*(2), 314-333.

Traugott, E. C. (1999). A Historical Overview of Complex Predicates. In L. J. Brinton & M. Akimoto (Ed.), *Collocational and Idiomatic Aspects of Composite Predicates in the History of English* (pp. 239–260). John Benjamins.

Tsou, B. K., & Yip, K. F. (2020). A corpus-based comparative study of light verbs in three Chinese speech communities. In *Proceedings of the 34th Pacific Asia conference on language, information and computation,* 302-311. https://aclanthology.org/2020.paclic-1.35

Tu, Y., & Roth, D. (2011). Learning English light verb constructions: contextual or statistical. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World,* 31-39. https://aclanthology.org/W11-0807

Vaidya, A., & Wittenberg, E. (2020). Productivity and argument sharing in Hindi light verb constructions. *Journal of South Asian Linguistics*, 11, 52-82.

Vaidya, A., Rambow, O., & Palmer, M. (2019). Syntactic composition and selectional preferences in Hindi light verb constructions. *Linguistic Issues in Language Technology*, 17. https://aclanthology.org/2019.lilt-17.1

VanValin, R. D., & La Polla, R. J. (1997). *Syntax. Structure, Meaning and Function.* Cambridge University Press.

Verdecchia, M. (2021). Nombres escuetos en contextos no verídicos. *Verba: Anuario Galego de Filoloxía*, 48. https://doi.org/10.15304/verba.48.6109

Vincze, V., Nagy T, I., & Farkas, R. (2013). Identifying English and Hungarian light verb constructions: A contrastive approach. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, 255–261.

Vilchez Vivancos, F., & Sarmiento González, R. (2011). *Manual de lenguaje jurídico-administrativo*. Librería-Editorial Dykinson.

Wheeler, M., Yates, A., & Dols, N. (1999). *Catalan: A Comprehensive Grammar*. Routledge.

Wierzbicka, A. (1982). Why Can You Have a Drink When You Can't *Have an Eat?. *Language*, *58*(4), 753- 799.

Wiltschko, M. (1998). On the Syntax and Semantics of (Relative) Pronouns and Determiners. *The Journal of Comparative Germanic Linguistics,* 2, 143–181. https://doi.org/10.1023/A:1009719229992

Winhart, H. (2001). *Funktionsverbgefüge im Deutschen: Zur Verbindung von Verben und Nominalisierungen* [Doctoral dissertation]. Universität Tübingen.

Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R. & Kuperberg, G. (2014). The difference between "giving a rose" and "giving a kiss": sustained neural activity to the light verb construction. *Journal of Memory and Language*, 73, 31-42. https://doi.org/10.1016/j.jml.2014.02.002

Wittenberg, E., & Piñango, M. M. (2011). Processing light verb constructions. *The mental Lexicon, 6*(3), 393-413. https://doi.org/10.1075/ml.6.3.03wit

Wittenberg, E., Khan, M., & Snedeker, J. (2017). Investigating Thematic Roles through Implicit Learning: Evidence from Light Verb Constructions. *Front. Psychol.* 8, 1089. https://doi.org/10.3389/fpsyg.2017.01089

Wittenberg, E., & Snedeker, J. (2014). It takes two to kiss, but does it take three to give a kiss? Categorization based on thematic roles. *Language, Cognition and Neuroscience*, *29*(5), 635-641. https://doi.org/10.1080/01690965.2013.831918

Wood, J. (2011). Icelandic let-causatives and case. *Working papers in Scandinavian syntax*, 87, 1-52.

Xu, H., Jiang, M., Lin, J., & Huang, C. R. (2022). Light verb variations and varieties of Mandarin Chinese: Comparable corpus driven approaches to grammatical variations. *Corpus Linguistics and Linguistic Theory*, *18*(1), 145-173. https://doi.org/10.1515/cllt-2019-0049

Yamashita, J., & Jiang, N. A. N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *Tesol Quarterly, 44*(4), 647-668. https://doi.org/10.5054/tq.2010.235998

Zdrojewski, P., & Sánchez, L. (2014). Variation in accusative clitic doubling across three Spanish dialects. *Lingua*, 151, 162-176. https://doi.org/10.1016/j.lingua.2014.08.003

Zipf, G. K. (1949). *Human Behaviour and the Principle of Least Effort*. Addison Wesley Publishing.

Appendix 1. Database

Open access at: https://osf.io/nbje6/?view_only=57069d42f0dc48deb3c6bb25840fbe4e

Appendix 2. Distribution of obligation and optionality of the determiner

Table 1-4 presents the distribution of the determiner in the LVCs under analysis for each language.

ENGLISH LVCs	Optional DET	Obligatory DET
give	Advice, example, notice, chase	Speech, answer, try, hug, smile, kiss, nod, push, sigh
take	Advantage, action, breath, risk	Step, picture, walk, bite, nap, bath, shower, hike, drink, stroll, swim
make	Progress, profit, adjustment, mention	Decision, mistake, choice, call, claim, contribution, payment, announcement, assumption, accusation
Total	12/42 (28.57%)	30/42 (71.43%)

Table 1. Distribution of obligation and optionality of a determiner in English LVCs

GERMAN LVCs	Optional DET	Obligatory DET
geben	Antwort, Erlaubnis, Rat, Beschreibung, Unterricht, Versprechen, Warnung	Hinweis, Kuss
nehmen	Abschied, Abstand, Rücksicht, Einsicht, Einfluss, Anteil, Platz, Anstoß	
machen	Hoffnung, Anfang, Sport, Eindruck, Gedanke, Sorge, Angabe, Erfahrung	Fehler, Foto, Ausführung, Beobachtung
Total	23/29 (79.31%)	6/29 (20.60%)

Table 2. Distribution of obligation and optionality of a determiner in German LVCs

CATALAN LVCs	Optional DET	Obligatory DET
donar	Suport, resposta, consell, permís, empenta, conferència, ajuda, explicació	Cop, volta, definició, bufetada, pallissa, bes/ada
prendre	Impuls, bany, represàlia, resolució, consciència, nota, possessió, partit, iniciativa, precaució, distància, determinació	Decisió, molèstia
fer	referència, esforç, feina, ús, anàlisi, dibuix, viatge, visita, fotografia	Pregunta, petó, abraçada, aclariment
Total	29/41 (70.73%)	12/41 (29.27%)

354

Table 3. Distribution of obligation and optionality of a determiner in Catalan LVCs

SPANISH LVCs	Optional DET	Obligatory DET
dar	Autorización, ayuda, vuelta, paso, respuesta, clase, cambio, consejo, instrucción, permiso	Beso, golpe, salto, giro, bofetada
tomar	Represalia, nota, conciencia, iniciativa, precaución, acuerdo, distancia, determinación, descanso, baño	Decisión, fotografía, atajo
hacer	Carrera, énfasis, broma, colección, esfuerzo, referencia, uso, ejercicio, análisis, investigación, declaración	Pregunta, viaje
Total	31/41 (75.61%)	10/41 (24.39%)

Table 4. Distribution of obligation and optionality of a determiner in Spanish LVCs

Appendix 3. List of publications derived from the thesis

The work developed in this dissertation has given place to the following papers:

A. Related publications

Alvarez-Morera, Georgina (to appear). Light Verb Constructions at two ends of a continuum. In M. Bargalló Escrivà (Ed.), *Recerca en Humanitats 2022*. Publicacions Universitat Rovira i Virgili.

Alvarez-Morera, Georgina (2023). L'ús dels verbs de suport fer i donar en el català actual. In J. Marqués Messeguer & E. Aguilar Miró (Eds.), *(De) bat a bat: obertures i cruïlles en estudis recents sobre literatura i llengua catalanes.* Presses universitaires de Perpignan. https://doi.org/10.4000/books.pupvd.41594

Alvarez-Morera, Georgina (2022). Light Verb Constructions in English and Spanish: State of the Art. In M. Bargalló Escrivà (Ed.), *Recerca en Humanitats 2021.* Publicacions Universitat Rovira i Virgili, 9-25. https://doi.org/10.17345/9788413650111

B. Unrelated publications

Fullana, Celia; Alvarez-Morera, Georgina & Oltra-Massuet, Isabel (2022). Unergative Cognate Structures across English and Romance. In F. Gallardo del Puerto, M-C. Camus Camus & J.A. González-López (Eds.), *Moving beyond the Pandemic: English and American Studies in Spain.* Ediciones Universidad de Cantabria, 65-77. https://doi.org/10.22429/Euc2022.034

Gómez Seibane, Sara & Alvarez-Morera, Georgina (2022). El marcado diferencial de objeto en el español del siglo XIX en Cataluña: registro y contacto de lenguas. *Revista de Filología Española*, 102(1), 87-109. https://doi.org/10.3989/rfe.2022.004

UNIVERSITAT
ROVIRA i VIRGILI