

Imploring you to investigate more languages: A corpus-based contrastive study of Ukrainian and English causative constructions

Iryna Karamysheva

Lviv Polytechnic National University

iryna.d.karamysheva@lpnu.ua

Jakob Horsch

Slovak Academy of Sciences

jakob.horsch@korpus.sk

Abstract

This study provides a corpus-based contrastive analysis of causative constructions in English and Ukrainian of the type S1 FORCES/CAUSES P1 [S2 P2/V_{inf}], which are simultaneously secondary predication constructions. Defining ‘causative construction’ as a form-function pairing where a primary subject (S1) exerts a causal force on a secondary subject (S2) to perform an action (P2), the research moves beyond raw frequency counts to explore the interaction between lexical fillers and constructional slots. Our corpus study is based on 247,737 tokens from COCA and 221,132 tokens from GRAC. Using Simple Collexeme Analysis and employing log-odds ratio as an association measure, the study addresses two questions: (1) identifying the most prototypical Complement Taking Predicates (P1) in each language, and (2) determining how the analytic character of English vs the synthetic character of Ukrainian shapes a ‘causative continuum’ via a network of such constructions in contrasted languages.

The findings reveal that English exhibits decentralized prototypicality, i.e. functional leaders coexist with semantically specialized, high-association verbs. Ukrainian, however, shows unified prototypicality with specific lexemes dominating frequency and association strength. Additionally, we find significant cross-linguistic variation: while English uses an objective case for the secondary subject (S2), reflecting the Latin term ‘Accusativus cum Infinitivo’, Ukrainian remains sensitive to case. The study finds that Ukrainian uses case marking (Accusative vs. Dative) as a structural diagnostic for ‘directness’, where high-force coercive verbs govern the Accusative, while persuasive verbs tend toward the Dative, underscoring that a language’s morpho-logical character can influence its semantic net-work of causation.

Keywords: causative construction, English, Ukrainian, collexeme analysis, corpus study

1 Introduction

In this corpus-based contrastive analysis, we investigate non-finite complement constructions with a causative meaning. Widely studied in English, they have been described variously, including ‘Accusativus cum Infinitivo’, ‘object control constructions’, ‘equi-type structures’, ‘small clauses’, and ‘periphrastic/analytical causatives’. Causative constructions encode one event as the argument of a second event, hence they have also been called ‘secondary predication constructions’. Consider the examples in English (1) and Ukrainian (2), which illustrate the syntactic roles involved:

(1) *When we_{S1} forced_{P1} [him_{S2} to retire_{P2}] [...]*

(Family Guy, 2011)

(2) *Вони знають, що ніхто_{S1} не змусить_{P1} [їх_{S2} повертати_{P2} боргу_{complement}]*

‘They know that nobody will force them to repay their debts.’ (<https://lb.ua/>, 2020)

As the examples (1-2) show, there is a primary predicate (P1) and a secondary predicate (P2) as well as a primary subject (S1), and a secondary subject (S2). This results in two predications. The secondary predication is expressed by an objective-case pronoun or common-case noun in English and an accusative/dative-case pronoun/ noun (Ukrainian) – S2 – in combination with an infinitive (P2). This predication becomes the argument of P1 as a non-finite complement.

In terms of semantics, S1 expresses a causal force/agent while S2 represents a patient that is impelled by force of reason to perform an act P2/V_{inf}. The overall meaning of the construction can thus be summarized as follows: [S1 FORCES (P1) S2 P2 /V_{inf}]. It is noteworthy that the causee S2 simultaneously performs two roles. First, it is a patient under the causal force from agent S1; second, it acts as an agent of the action expressed by P2/V_{inf}. Thus, S2, which can be regarded an object of P1, is simultaneously the logical subject of P2. In the literature, the combination of S2+P2

has been treated as one argument, expressed by the non-finite complement construction (Noonan, 2007: 53; Croft, 2022: 551-558). Hence the predicates (P1) are referred to as ‘Complement Taking Predicates’ (CTPs).

There is a large body of research on the English causative construction, ranging from studies against the background of mainstream generative grammar (e.g., Comrie, 1976) to non-transformational approaches (e.g., Kim and Michaelis, 2020: 168 ff.). Recently, the causative construction has also attracted attention from constructionist research, which traditionally relies on corpus data (e.g., Egan, 2008; Gilquin, 2010). These studies have investigated periphrastic causative constructions using sophisticated statistical methods that are collectively referred to as collocation analysis. Developed by Stefanowitsch and Gries (2003; 2008), this method is used to measure the degree of interaction between words and constructions (Gilquin and Gries, 2026). For example, Gilquin (2006) employed multiple distinctive collexeme analyses to investigate the non-finite verb slots of ten English periphrastic causative constructions such as [X CAUSE Y _{Vto-inf}] (Gilquin, 2010, 2024). Demonstrating its feasibility for contrastive studies, Gilquin also applied collexeme analysis by comparing the infinitive verb slots of the English causative construction [X MAKE Y _{Vinf}] to its French equivalent [X FAIRE _{Vinf} Y] (2015, 2024).

In the following, we present a contrastive corpus study of the English causative construction and its Ukrainian counterpart, relying on collocation analysis to uncover formal differences and similarities. While the Ukrainian causative construction remains largely unexplored, some recent corpus studies have provided a first glimpse (Karamysheva, 2024, 2025). These studies, while limited in scope, have revealed that Ukrainian differs from English in some respects. For example, in English the non-finite V slot P2 can be filled not only by infinitives with and without to, but also with the Present Participle and Past Participle (Gilquin, 2006; Karamysheva, 2025). In Ukrainian, however, the V slot can only be filled by an infinitive.

In the following, we restrict ourselves to the P1/CTP slot, presenting the first-ever contrastive corpus-based investigation of English and

Ukrainian causative constructions. Specifically, we will explore so-called causative verbs that fill the CTP slot. Our research questions are, thus, as follows:

1. Which CTPs are most prototypically used in the English and Ukrainian causative constructions?
2. Are there any differences between the two languages, and how can they be explained?

Following a “the more data, the better” approach (Hoffmann et al., 2019), we rely on large datasets extracted from the Corpus of Contemporary American English (COCA) (Davies, 2008) and the General Regionally Annotated Corpus of Ukrainian (GRAC) (Shvedova et al., 2017), following the consideration that both are large and representative. Assuming a constructionist approach that treats syntactic patterns that encode specific non-compositional meanings as form-meaning pairings in their own right (cf., e.g., Horsch and Boas, 2025: 16), we employ the term ‘causative construction’ in the following.

2 Methodology

2.1 Data preparation

First, we used synonymic expansion to compile lists of verbs that can be used as CTPs in the English and Ukrainian causative constructions. For English, we additionally relied on a variety of sources (e.g., Quirk et al., 1985; Noonan, 2007; Egan, 2008). The list of Ukrainian causative CTPs was compiled using the ‘synonym function’ of the online resource slovnkyk.me¹, which is based on authoritative dictionaries of the Ukrainian language. We had to rely on this procedure because to our knowledge, there are no ready-made lists of CTPs that trigger this type of non-finite complement construction in Ukrainian. For both languages, we cross-checked the corpus data whether the selected verbs fit to the syntactic pattern N₁ V N₂ non-f V (N₃).

Our choice of corpora follows from the fact that Ukrainian, unlike English, is a low-resource language, with a very limited choice of corpora (as noted by Shvedova and Lukashevskyi, 2024: 14). Thus, while we acknowledge that a balanced, bidirectional corpus would be ideal for providing a *tertium comparationis* (TC) (as suggested by Hasselgård 2020), we settled on COCA and

¹ <https://slovnkyk.me/dict/synonyms> (last accessed 13.03.2026)

GRAC, following Karamysheva, who has noted that “two separate corpora [...] can be used to harvest data for contrastive analysis, provided that there is a correct choice of [TC], capturing a certain pattern, and reflected in the query built” (2024: 1). Henceforth we shall, therefore, use the constructionist notion of ‘construction’ (see above) to serve as a TC, allowing us to compare English and Ukrainian regarding CTPs.

Our queries differed slightly between the corpora: For COCA, we used the features offered by the online interface at english-corpora.org; for GRAC we used CQL. The search of causative constructions with specific CTPs presupposed two queries for each CTP, since the non-finite complement contains a causee/patient (S2) expressed either by a pronoun in the objective case or a noun in the common case (in English) or pronoun/noun either in the accusative or dative case, depending on the CTP (in Ukrainian). Consider the queries and the examples of constructions obtained (3-6):

(3) CAUSE PRON TO _v?i

Every talk has caused me to think about some topic a new way.

(BLOG <http://lfb.org/today/the-center-of-the-conspiracy/>, 2012)

(4) ENCOURAGE NOUN TO _v?i

School librarians encourage students to consider diverse and global perspectives

(ACAD: Knowledge Quest, 2015)

(5) [lemma="наказувати"] [tag="*.pron.*"] [tag="*.inf.*"]

[...] *поліція наказує їм залишатися на місці.*

‘[...] the police order them to stay in place.’

(<https://www.eurointegration.com.ua/>, 2019)

(6) [lemma="радити"] [tag="*.noun.*" & tag!="*.pron.*"] [tag="*.inf.*"]²

Через це експерти з кібербезпеки радять людям переходити на інші застосунки.

‘Because of this, cybersecurity experts are advising people to switch to other apps.’

(<https://nv.ua/>, 2021)

In total, we extracted 27 English and 23 Ukrainian CTPs with semantics typically associated with the causative construction. In each case, the token frequencies of both queries were

summed up to provide the general frequency of the construction per corpus (this value is required for collocation analysis, see section 2.3 below). Moreover, to investigate whether the ‘causative continuum’ differs between the two languages (cf. Talmy, 1988; Comrie, 1989; Shibatani and Pardeshi, 2002), to the effect that the ‘force’ of causation can be more direct or less direct, as reflected by the semantic spectrum of causative verbs, we subdivided our lists of causatives into two groups: 1) ‘coercive’ with a more direct causation impact and 2) ‘persuasive’ with a milder form of causation, expressed by requests, advice or persuasion.

The list of English coercive verbs includes the following 11 CTPs, in order of descending token frequency: *make* (88,541), *force* (21,382), *cause* (14,652), *get* (9,379), *have* (5,370), *bring* (3,422), *compel* (1,915), *drive* (1,580), *command* (749), *press* (428), *oblige* (403). The English persuasive type includes the following 16 CTPs: *ask* (48,523), *encourage* (15,936), *lead* (10,993), *urge* (8,137), *beg* (3,047), *prompt* (2,910), *advise* (2,881), *persuade* (2,874), *challenge* (2,162), *dare* (1,137), *implore* (445), *influence* (347), *recommend* (187), *stimulate* (187), *beseech* (137), *plead* (013).

The list of Ukrainian coercive verbs includes 7 CTPs: *примусити|змусити* ‘make, force’ (52,473); *казати* ‘force, oblige’ (2,830); *велити* ‘command, order’ (2,580); *наказувати* ‘order’ (2,145); *заставляти* ‘force’ (678); *сильувати* ‘force’ (239); *приневолювати* ‘force’ (173). The list of Ukrainian persuasive verbs is longer and includes 16 CTPs: *попросити|просити* ‘ask’ (86,354); *пропонувати|запропонувати* ‘advise’ (38,889); *порекомендувати|рекомендувати* ‘recommend’ (9,315); *радити|порадити* ‘advise’ (8,005); *спонукати* ‘urge’ (7,931); *умовляти|вмоляти* ‘persuade’ (2,052); *заохочувати* ‘encourage’ (1,753); *стимулювати* ‘encourage’ (1,670); *благати* ‘beg, beseech’ (1,597); *переконувати* ‘persuade’ (1,474); *підбурювати* ‘urge, encourage’ (331); *підбурювати* ‘persuade, coerce’ (212); *схилити* ‘persuade, incline’ (186); *підбивати* ‘persuade, coerce’ (130); *заклинати* ‘beg, beseech’ (58); *раяти|нараяти* ‘advise, recommend’ (57).

Note that some Ukrainian CTPs are treated as pairs, following two considerations: First, some CTPs (e.g.

² Note that the query for Ukrainian constructions with pronouns as S2 (Example (5), [lemma="наказувати"] [tag="*.pron.*"] [tag="*.inf.*"]) yielded clean results, whereas the query with the noun tag, e.g. [lemma="наказувати"] [tag="*.noun.*"]

[tag="*.inf.*"] returned sentences intermixed with pronouns, which called for the manual sorting out of such cases. Therefore, a more specialized query, cf. example (6), was required.

nonpocumu|npocumu ‘ask’) are aspectual pairs of the same lexeme; second, some are allomorphic variants (e.g. *npumycumu|zmycumu* ‘make, force’), which may be additionally treated as interchangeable synonyms.

2.2 Data extraction

To address the research questions, we annotated data from two corpora, COCA 2019³ and GRAC v. 19a⁴. Table 1 provides an overview of the corpora and how many tokens were retrieved.

Corpus	Corpus size (tokens)	No. of hits (tokens)
COCA	1,001,610,938	247,737
GRAC	2,023,921,236	221,132

Table 1: Overview of corpora and tokens

The extracted data were saved in spreadsheets and annotated in *Libre Office Calc* (The Document Foundation, 2025) for the following variables:

- WORD (levels: e.g. *make, have...*)
- FREQUENCY WORD IN CORPUS
- FREQUENCY WORD IN CONSTRUCTION

These variables are required as input for collexeme analysis, which we discuss in the next section.

2.3 Data analysis: Collexeme analysis

We use statistical tests to explore semantic tendencies in the CTP slot of the English and Ukrainian causative constructions. Specifically, we employ simple collexeme analysis (Stefanowitsch and Gries, 2003; Stefanowitsch and Gries, 2008; Stefanowitsch 2013), which “identifies association patterns [...] between lexical items and syntactic constructions” (Hartmann and Ungerer, 2023: 608). Put succinctly, collexeme analysis “identifies lexemes that occur with above-chance frequency in a construction”; these are referred to as “attracted collexemes” (Hartmann and Ungerer, 2023: 608). Simple collexeme analysis works by calculating expected frequencies on the basis of contingency tables and applying “inferential statistical tests [to determine] the statistical significance of the observed distribution” (Stefanowitsch, 2014: 221).

Simple collexeme analysis outputs a list that shows “which verbs are typical for a given

construction” (Stefanowitsch, 2014: 222). Following the assumption that “the verbs most typically occurring with a given construction reflect its meaning”, this allows us to draw conclusions about the construction itself (Stefanowitsch, 2014: 222), allowing us to address the research questions. CTPs that turn out to be ‘attracted’ to the causative construction are referred to as ‘collexemes’.

To carry out collexeme analysis, we employed Gries’ (2024) *Coll.analysis_4.1* script for R (R Core Team, 2020). As association measure (AM)⁵, we rely on the log odds (henceforth referred to as ‘log-odds’), which has been described as the “gold standard [...] for association” (Gries, 2022: 27). The main advantage of log-odds is that, unlike other commonly used measures (e.g. G^2 or t), it “behaves exactly like a measure that was designed to control frequency” (Gries, 2022: 26). In fact, apart from a results table in which collexemes are ranked by log-odds, *Coll.analysis_4.1* conveniently also outputs a graph plotting log-odds (the AM) against the y-axis and frequency against the x-axis (as logged co-occurrence frequencies).

The importance of distinguishing between AM and frequency will become more apparent when we turn to the results in section 3. For now, suffice it to say that corpus data can and do show all four possible combinations (as noted by Gries, 2022: 30):

- high frequency and high association
- high frequency and low association
- low frequency and high association
- low frequency and low association

Obviously, the fact that all combinations are possible has implications for interpreting corpus data, because it demonstrates that frequency or AMs by themselves do not necessarily correlate and therefore cannot be used for drawing valid conclusions about the relationship between a word and the construction in whose context it is examined (Gries, 2022). Rather, AMs and frequency should complement each other, to the effect that the latter is used to contextualize the former (Gries, 2022: 13).

³ <https://www.english-corpora.org/coca/>

⁴ <https://uacorpus.org/>

⁵ We acknowledge that recently, there has been a trend towards “tupleization” (Gries, 2026), i.e., an “emerging consensus among users of collocation-based methods to

ideally draw on a combination of different measures” (Hartmann and Ungerer, 2023: 609). Due to the space limitations of this publication, we have decided to restrict ourselves to log-odds.

3 Results

3.1 English

First, we present the results for English. Table 2 shows the results of simple collexeme analysis, including frequency counts. We see that of all CTPs, *implore* and *encourage* (log-odds: 7.09) are semantically most strongly attracted to the English causative construction, followed by *urge* (log-odds: 7.04). At the bottom of Table 1, least strongly attracted, we find *have* (log-odds: 0.58), *plead* (log-odds: 1.06) and *press* (log-odds: 2.31).

Word	Freq. in cxn	Freq. not in cxn	Rel.	Log- odds
<i>implore</i>	445	1497	attr.	7.09
<i>encourage</i>	15936	57316	attr.	7.09
<i>urge</i>	8137	29903	attr.	7.04
<i>persuade</i>	2874	14355	attr.	6.71
<i>compel</i>	1915	12227	attr.	6.46
<i>beseech</i>	137	876	attr.	6.45
<i>beg</i>	3047	22070	attr.	6.34
<i>prompt</i>	2910	21962	attr.	6.30
<i>advise</i>	2881	24699	attr.	6.17
<i>ask</i>	48523	628216	attr.	5.96
<i>force</i>	21382	296333	attr.	5.77
<i>oblige</i>	403	6261	attr.	5.56
<i>make</i>	88541	2207658	attr.	5.53
<i>cause</i>	14652	289988	attr.	5.38
<i>dare</i>	1137	25743	attr.	5.19
<i>lead</i>	10993	334528	attr.	4.93
<i>stimulate</i>	187	10726	attr.	4.26
<i>challenge</i>	2162	145204	attr.	4.11
<i>command</i>	749	50026	attr.	4.11
<i>bring</i>	3422	436031	attr.	3.47
<i>drive</i>	1580	252442	attr.	3.24
<i>influence</i>	347	104479	attr.	2.60
<i>recommend</i>	187	56751	attr.	2.59
<i>get</i>	9379	3341902	attr.	2.46
<i>press</i>	428	171233	attr.	2.31
<i>plead</i>	13	18116	attr.	1.06
<i>have</i>	5370	12310549	attr.	0.58

Table 2: Collexeme analysis results, English

Figure 1 shows that AM and frequency do not necessarily correlate, and that a complementary approach is therefore warranted (see discussion in section 2.2). For example, with 445 tokens, *implore* has a considerably lower frequency than *have* (5,370 tokens). Nevertheless, with a log-odds of

7.09, *implore* is much more strongly associated with the causative construction than *have* (log-odds: 0.58). Returning to the plot in Figure 1, this notion becomes clear. Recall that log-odds is plotted against the y-axis and logged co-occurrence frequency against the x-axis. This means that generally speaking, the more to the top right of the semantic space we find a CTP, the more prototypical it can be considered for the construction. Conversely, the more to the bottom left of the plot a word is found, the less prototypical it is for the construction. Thus, while *implore* has the highest AM (log odds: 7.09), it actually has a considerably lower frequency (445 tokens), making it less prototypical than e.g. *ask* (48,523 tokens, log-odds: 5.96) and *make* (88,541 tokens, log-odds: 5.53).

More prototypical CTPs include *encourage* (15,936 tokens, log-odds: 7.09), *force* (21,382 tokens, log-odds: 5.77) and *cause* (14,652 tokens, log-odds: 4.38). Conversely, *plead*, with a low frequency (13 tokens) and a low AM (log-odds: 1.06) may be considered the least prototypical. Finally, as discussed in section 2, there is also the possibility of high frequency combined with low association, such as *have* (5,370 tokens, log odds: 0.58). The plot also shows what might be termed clusters⁶ of CTPs, for example *persuade*, *beg*, *prompt*, *advise*.

Finally, when considering the ‘big picture’, the plot reveals a rather heterogenous distribution, with a considerable spread across the y- and x-axes. Thus, there is quite a large difference between the lowest (*have*, 0.58) and highest (*implore*, 7.09) log-odds, making for a range of 6.51. This stands in contrast with the Ukrainian results, a point to which we return in section 3.2 below. While our range of causative verbs differs from that used in previous research, making comparisons difficult, we would like to note Gilquin’s (2010) corpus study of English periphrastic causative constructions (with e.g. *make*, *cause* and *have*), which employed a related statistical method, multiple distinctive collexeme analysis. Gilquin found that the pattern [X *make* Y V_{inf}] turns out to be the most flexible, attracting the highest number of distinctive features in comparison to other verbs.

⁶ By ‘clusters’ we mean CTPs in close proximity to each other on the plot, i.e., with similar frequencies (x axis) and log-odds values (y axis).

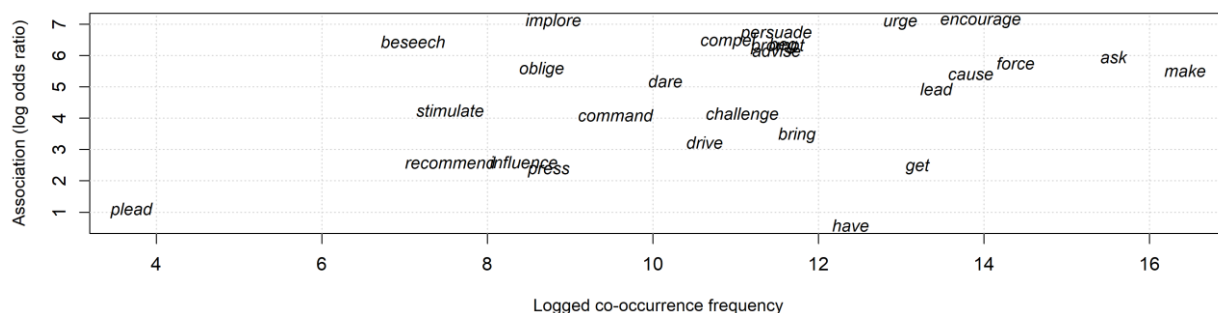


Figure 1: Plot of CTPs in the English causative construction, COCA

She speculates that “people associate [X *make* Y V_{inf}] [as] the most frequent and flexible periphrastic causative construction, with inductive causation because it is a meaning that is somehow more salient in people’s minds than the other types of causation” (Gilquin, 2010: 142). In this context, it is intriguing that *make* is, by far, the most frequent verb in our data, with 88,541 tokens (cf. Table 2). However, it does not exhibit the highest AM (log-odds: 5.53), occupying more of a middle position, an observation we believe can be explained with its polysemous nature.

3.2 Ukrainian

Turning to the Ukrainian results (Table 3 and Figure 2), we see some striking differences to the English results from section 3.1. When considering the plot, we see a more homogenous distribution, with more spread along the *x*-axis but not as much over the *y*-axis. This is also evident from the fact that in the Ukrainian data, the log-odds values (plotted against the *y*-axis) range from 3.17 (*казати* ‘say’ used in the meaning ‘to order’) to 9.10 (*примусити*|*змусити* ‘force, compel’), a difference of 5.93.

However, the lower bound in the Ukrainian data is an outlier (*казати*), without which the range would be just 3.68. What this suggests is that despite a spread in terms of frequency that is comparable to English, the attraction of CTPs to the causative construction in Ukrainian is more uniform, indicating less variability. That is, frequency does not appear to interact much with the AM, which could be interpreted as individual CTPs being rather equal in terms of how they are used in the Ukrainian causative construction. Turning to the individual CTPs, we see that with a high frequency and a high AM, *примусити*|*змусити* ‘force, compel’ (52,473 tokens, log-odds: 9.10) and *попросити*|*просити* ‘ask, request’ (86,354 tokens, log-odds: 7.50) are most prototypical for the Ukrainian causative construction.

Word	Freq. in cxn	Freq. not in cxn	Rel.	Log-odds
<i>примусити</i> <i>змусити</i>	52473	70105	attr.	9.10
<i>приневоляти</i>	173	524	attr.	8.01
<i>спонукати</i>	7931	27360	attr.	7.92
<i>умовляти</i> <i>вмовляти</i>	2052	9146	attr.	7.64
<i>заставляти</i>	678	3028	attr.	7.63
<i>попросити</i> <i>просити</i>	86354	720449	attr.	7.50
<i>веліти</i>	2580	13606	attr.	7.47
<i>наказувати</i>	2145	15668	attr.	7.14
<i>силувати</i>	239	1770	attr.	7.12
<i>заохочувати</i>	1753	13453	attr.	7.09
<i>порекомендувати</i> <i>екомендувати</i>	9315	82904	attr.	6.98
<i>пропонувати</i> <i>запропонувати</i>	38889	559252	attr.	6.65
<i>благати</i>	1597	20404	attr.	6.58
<i>раяти</i> <i>нараяти</i>	57	726	attr.	6.58
<i>порадити</i> <i>радити</i>	8005	107603	attr.	6.56
<i>підбурювати</i>	212	3138	attr.	6.43
<i>стимулювати</i>	1670	27316	attr.	6.33
<i>підштовхувати</i>	331	7022	attr.	6.07
<i>заклинати</i>	58	1307	attr.	6.01
<i>переконувати</i>	1474	39078	attr.	5.85
<i>підбивати</i>	130	4263	attr.	5.63
<i>схиляти</i>	186	7522	attr.	5.42
<i>казати</i>	2830	1105235	attr.	3.17

Table 3: Collexeme analysis results, Ukrainian

With low frequencies and (relatively) low AMs, the least prototypical CTPs are *раяти*|*нараяти* ‘advise, recommend’ (57 tokens, log-odds: 6.58) and *заклинати* ‘beg, beseech’ (58 tokens, log-odds: 6.01). These results may be due to the fact that *раяти*|*нараяти* ‘advise, recommend’ is considered rather dialectal and antiquated, whereas *заклинати* ‘beg, beseech’ is more characteristic of literary style than everyday conversation. Again, note that these CTPs are found in the middle of the table, making them rather inconspicuous. Only

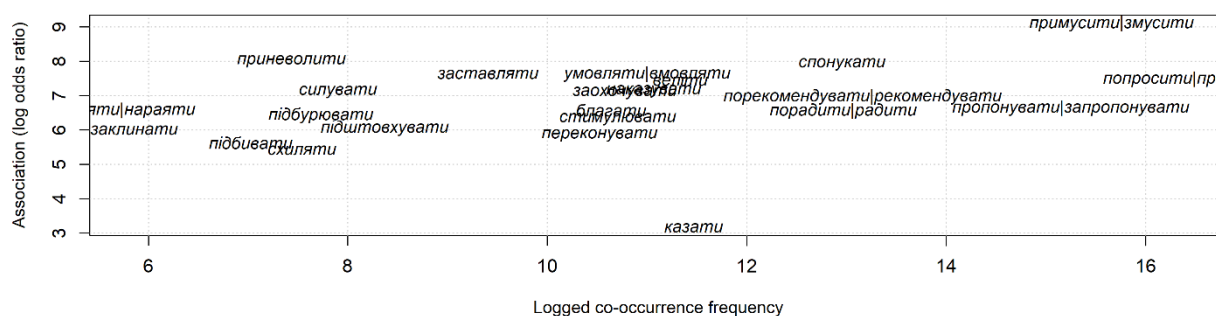


Figure 2: Plot of CTPs in the Ukrainian causative construction, GRAC

when taking into account their low frequencies can we appreciate their comparatively ‘weak’ prototypicality, underscoring the importance of the complementary approach, i.e., contextualizing association measures with frequencies (see section 2.2).

4 Discussion

The results of our simple collexeme analyses reveal similarities and differences between English and Ukrainian, with the main difference being that causative verbs prioritize different types of CTPs as prototypes. Thus, regarding research question 1, we conclude that in English the verbs *ask* (log-odds: 5.96) and *make* (log-odds: 5.53) are the most prototypical, with high corpus frequencies and high association measures (AM). By both measures, they clearly distinguish themselves from the other verbs. Interestingly, English features many verbs (e.g. *encourage* and *force*) with a vastly higher AM than the more generic *have*. In fact, the low AM of *have* (log-odds: 0.58) indicates what we call a ‘polysemous drifter’ – with a high overall frequency in the language and not specifically dedicated to the causative construction. In Ukrainian, in contrast, the prototypical ‘core’ is more compact. *Примусити/змусити* ‘force, compel’ and *nonпросити/просити* ‘ask, request’ emerge as the verbs most attracted to the construction. The high AM (e.g. *примусити|змусити* ‘force, compel’; log odds: 9.10) suggests that when a Ukrainian speaker produces this construction, s/he is statistically much more likely to rely on a ‘highly coercive’ causative verb than an English speaker.

Our results also suggest that the ‘causative continuum’ – the scale from direct coercive manipulation to indirect persuasion – grades differently in English and Ukrainian. English displays heterogeneity and a high degree of lexical specialization; the wide range of log-odds (6.51)

and the ‘semantic clusters’ (e.g. around *persuade* and *beg*, cf. Figure 1) indicate a rather fragmented continuum. Ukrainian, on the other hand, presents a more homogenous distribution. Excluding the outlier *казати* ‘say’, the AM range is significantly more restricted than in English. Moreover, the consistently high AM values in Ukrainian suggest that the construction ‘attracts’ its verbs/CTPs more strongly, leading to a more uniform attraction across the investigated lexicon. Regarding the second research question, we can, thus, conclude that differences outweigh similarities in the distributional behavior, even if the semantic intent remains the same with 16 CTPs in each language of the persuasive subtype.

On a final note, our Ukrainian data shows that some CTPs use dative case on the secondary subject S2, warranting the label *Dativus cum Infinitivo* (DcI). This concerns 3 out of 7 coercive CTPs (*казати* ‘order’, *велити* ‘command/order’, *наказувати* ‘order’) and 4 out of 16 persuasive CTPs (*пропонувати|запропонувати* ‘offer’, *порекомендувати|рекомендувати* ‘recommend’, *радити* ‘advise’, *ряти|нараяти* ‘advise, recommend’). It, thus, appears that in Ukrainian, in contrast with English, morphology/case is sensitive to verb semantics (‘force dynamics’ in Talmy’s (1988) sense.). This is in line with previous research in Slavonic languages on object control constructions that prefer dative case (e.g., Dziwirek, 1998). In fact, there is a substantial body of research devoted to causation in different languages, impossible to cite in its entirety without going beyond the scope of this paper. These include studies with a focus on the typology of causatives, including their form, syntax and meaning, e.g. Dixon, who remarks that “languages differ a great deal in the syntax of their causatives and in the specific meanings attached to them” (2000: 33). Shibatani and Pardeshi note that “when a language allows alternative marking of the causee nominal, the accusative version generally conveys sociative causation, whereas the dative or

other oblique marking signals indirect causation” (2002: 99-100). Since it is the Ukrainian language that is understudied regarding typology, we consider the results obtained in this study of great value.

In this context, it can be stated that in Ukrainian causative constructions the accusative case marks direct impact, while dative marks interpersonal communication. Both retain ‘complex transitivity’ because they both follow this pattern, but they use case as a precision tool to define the relationship between the primary S1 and secondary S2 subjects. Ukrainian CTPs of the ‘accusative cluster’ (e.g. *примусити/змусити* ‘make, force’, *заставляти* ‘force’ and *сильувати* ‘compel’) express high coercion/ direct manipulation. They are ‘true’ causatives in the sense that the causer acts directly on the secondary subject. The high token count for *змусити* ‘force, compel’ suggests this is the prototypical periphrastic causative in Ukrainian, much like *make* in English. The ‘dative cluster’ of coercive verbs can be regarded as less directive/more social (i.e., indirect manipulation). CTPs like *казати* (here in the sense of ‘tell, command’), *велити* ‘command’ and *наказувати* ‘order’, encode by means of the dative case that the secondary subject is not just an object being pushed, but an addressee of a speech act who is expected to carry out the secondary predication action. In contrast, the most prototypical persuasive CTP *попросити/просяти* ‘ask’ takes the accusative, which suggests that ‘asking’ is still viewed as a direct social impact on the person, while the dative persuasives (e.g. *пропонувати* ‘offer/advise’, *порекомендувати* ‘recommend’ and *підказати* ‘advise’ indicate a weaker causal link where the secondary subject is a volitional addressee rather than a manipulated object.

5 Conclusion

In this study of the causative construction, we used simple collexeme analysis to analyze 27 causative CTPs in English and 23 causative CTPs in Ukrainian. The results show interesting similarities and differences between the two languages. Most prominently, English appears more heterogeneous in how verbs are used in the construction, whereas Ukrainian turns out to be quite homogeneous.

In more visual terms, English is best described as exhibiting lexical dispersion, with a loose agglomeration of various verbs with weak-to-moderate attraction. Ukrainian, in contrast, features a more uniform and specialized set of CTPs that are more fundamentally attracted to the causative construction. Even mid-to-low frequency verbs in Ukrainian show attraction levels that rival the very highest peaks in English, where we also observe a long tail of verbs with very low association (log-odds <3.0). In contrast, in Ukrainian even the less frequent verbs maintain association strengths that would place them near the very top of the English list. Regarding our differentiation between ‘coercive’ and ‘persuasive’ causative CTPs, we observe that in Ukrainian, case marking acts as a structural diagnostic for different levels of ‘directness’ and ‘interpersonal manipulation.’ From a methodological perspective, our study underscores the importance of using an association measure such as the log-odds in addition to raw corpus frequencies (cf. section 2.3 and Gries, 2022: 30).

Being a pilot study, our findings are to be considered preliminary, with more extensive research with a focus on Ukrainian certainly being warranted. For example, registers will have to be taken into account, given the results for some verbs that are e.g. dialectal/ antiquated (see section 3.2). On a final note, due to limitations regarding time and resources, we were unable to discuss our results in the context of the stylistic/genre structure of the corpora⁷. However, we are convinced that taking into account the distribution of the studied constructions across different genres of the two corpora constitutes an interesting topic for further research.

Acknowledgments

We would like to thank two anonymous reviewers for their valuable suggestions on how to improve our manuscript. All remaining errors are our own.

Part of the work reported here was carried out in the context of the research project “Weight Effects in English and Slovak” (2024–26), generously funded by the European Union’s *NextGenerationEU* program through the Recovery and Resilience Plan for Slovakia (project no. 09I03-03-V04-00713), in the context of which one of the authors was employed.

⁷ We would like to express our thanks to an anonymous reviewer for pointing this out.

References

- Comrie, Bernard. 1976. The syntax of causative constructions: cross-language similarities and divergences. In M. Shibatani (ed.) *Syntax and Semantics, Vol. 6: The Grammar of Causative Constructions*. New-York, San Francisco, London: Academic Press, pages 261–312.
- Comrie, Bernard. 1989. *Language Universals and Linguistic Typology: Syntax and Morphology*. University of Chicago Press.
- Croft, William. 2022. *Morphosyntax: Constructions of the World's Languages*. Cambridge: Cambridge University Press.
- Davies, Mark. 2008. *The Corpus of Contemporary American English (COCA)*: 520 million words, 1990-present. Available at: <https://www.english-corpora.org/coca/>
- Dixon, Robert M.W. (2000). A typology of causatives: form, syntax and meaning. In R. M. W. Dixon and Alexandra Y. Aikhenvald (eds.) *Changing valency: case studies in transitivity*. Cambridge University Press, 2000, pages 30–83.
- Dziwirek, Katarzyna. 1998. Reduced Construction in Universal Grammar: Evidence from the Polish Object Control Construction, *Natural Language and Linguistic Theory*, 16: 53–59.
- Egan, Thomas. 2008. *Non-finite Complementation. A Usage-based Study of Infinitive and -ing Clauses in English*. Amsterdam-New York.
- Gilquin, Gaëtanelle. 2015. Contrastive collostructional analysis: Causative constructions in English and French. *Zeitschrift für Anglistik und Amerikanistik*, 63(3): 253–272.
- Gilquin, Gaëtanelle. 2010. *Corpus, Cognition and Causative Constructions*. Studies in Corpus Linguistics, V. 39, John Benjamins.
- Gilquin, Gaëtanelle. 2006. The verb slot in causative constructions. Finding the best fit', *Constructions*, SV1. Available at: <https://constructions.journals.hhu.de/article/view/445>
- Gilquin, Gaëtanelle. 2024. Transfer of collostructions: the case of causative constructions, *Corpus Linguistics and Linguistic Theory*, pages 1–21. Available at: <https://doi.org/10.1515/cllt-2024-0023>
- Gilquin, Gaëtanelle and Gries, Stefan. Th. 2026. Collostructional analysis. In H. Nesi and P. Milin (eds.) *International Encyclopedia of Language and Linguistics*, 3rd Edition. Elsevier.
- Gries, Stefan Th. 2022. What do (some of) our association measures measure (most)? Association? *Journal of Second Language Studies*, 5(1):1–33.
- Gries, Stefan Th. 2024. Coll.analysis 4.1. A script for R to compute perform collostructional analyses.
- Gries, Stefan Th. 2026. Tupleization. in H. Nesi and P. Milin (eds.) *International Encyclopedia of Language and Linguistics*. 3rd ed. Amsterdam: Elsevier.
- Hartmann, Stefan. and Ungerer, Tobias. (2023) “Attack of the snowclones: A corpus-based analysis of extravagant formulaic patterns,” *Journal of Linguistics*, 60(3): 599–634.
- Hasselgård, Hilde. 2020. Corpus-based contrastive studies: Beginnings, developments and directions, *Languages in Contrast*, 20(2): 184–208.
- Hoffman, Thomas, Horsch Jakob, and Brunner Thomas. 2019. The More Data, The Better: A Usage-based Account of the English Comparative Correlative Construction. *Cognitive Linguistics*, 30(1): 1–36.
- Horsch, Jakob and Boas, Hans C. (2025) Construction Grammar-Based Research on Slovak, in Slovakia: A Case Study and a Road Map. *Slovak Speech*, 90(2): 11–50. Available at: <https://doi.org/10.31577/slovrec.2025.2.2>
- Karamysheva, Iryna. 2024. Challenges of corpus-based contrastive studies: a case study of the English “make”-construction and its Ukrainian counterpart. *IEEE 19th International Conference on Computer Science and Information Technologies (CSIT)*, Lviv, Ukraine, pages 1–4.
- Karamysheva, Iryna. 2025. The choice of corpora for contrastive analysis: a case study of the English causative construction with “encourage” and its Ukrainian counterpart.

- Іноземна філологія [Foreign Philology]*. Vol. 138, pages 133–147).
- Kim, Jong-Bok and Michaelis, Laura A. 2020. *Syntactic Constructions in English*. Cambridge: Cambridge University Press.
- Noonan, Michael. 2007. Complementation. in T. Shopen (ed.), *Language Typology and Syntactic Description*, vol. II: Complex constructions, 2nd edition. Cambridge: Cambridge University Press, pages 52–150.
- Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, and Svartvik, Jan. 1985. *A comprehensive grammar of the English language*. Indexed by D. Crystal. London: Longman.
- R Core Team. 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>
- Shibatani, Masayoshi, and Pardeshi, Prashant. 2002. The causative continuum. In M. Shibatani (ed.), *The Grammar of Causation and Interpersonal Manipulation*. John Benjamins Publishing, pages 85–126.
- Shvedova, Maria and Lukashevskyi, Arsenii. 2024. Creating parallel corpora for Ukrainian: A German-Ukrainian parallel corpus (ParaRook||DE-UK). *Proceedings of the Third Ukrainian Natural Language Processing Workshop (UNLP) @ LREC-COLING*. Torino, Italia. ELRA and ICCL, pages 14–22.
- Shvedova, Maria *et al.* 2017. GRAC: General Regionally Annotated Corpus of Ukrainian. Kyiv, Lviv, Jena. Available at: uacorporus.org.
- Stefanowitsch, Anatol and Gries, Stefan Th. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics*, 8(2): 209–243. Available at: <https://doi.org/10.1075/ijcl.8.2.03ste>
- Stefanowitsch, Anatol and Gries, Stefan Th. 2008. Channel and constructional meaning: A collostructional case study. In G. Christiansen and R. Dirven (eds.) *Cognitive sociolinguistics: language variation, cultural models, social systems*. Berlin: Mouton de Gruyter (Cognitive linguistics research, 39), pages 129–152.
- Stefanowitsch, Anatol. 2013. Collostructional analysis. In T. Hoffmann and G. Trousdale (eds.) *The Oxford Handbook of Construction Grammar*. Oxford: Oxford UP, pages 290–306. Available at: <https://doi.org/10.1093/oxfordhb/9780195396683.013.0016>
- Stefanowitsch, Anatol. 2014. Collostructional analysis: A case study of the English into-causative. In T. Herbst, H.-J. Schmid, and S. Faulhaber (eds.) *Constructions Collocations Patterns*. Berlin, München, Boston: De Gruyter Mouton, pages 217–238.
- Talmy, Leonard. 1988. Force Dynamics in Language and Cognition. *Cognitive Science*, 12(1): 49–100.
- The Document Foundation. 2025. LibreOfficeCalc 25.8.4.2. Available at: <https://www.libreoffice.org/discover/calc/>