

Lin Ding, Kristel Van Goethem & Ferran Suñer Muñoz

One Hell of a Competition:

Register Distribution, Lexical Productivity, and Semantic Preferences of [*a HELL of a(n) N*] and [*a HELL(-)N*]

Abstract: This article focuses on the competition between the evaluative binominal construction [*a HELL of a/an N₂*] and the evaluative compound construction [*HELL(-)N₂*]. Drawing on data from the *English Web 2021* corpus and the English component of *OpenSubtitles2018*, the study investigates their distribution and productivity across registers, their lexical and semantic preferences, and the factors conditioning constructional choice. The results show that the phrasal construction predominates across registers, whereas the compound construction demonstrates greater realized and potential productivity. The two constructions also differ systematically in the head nouns they attract and in the semantic domains represented by those nouns. Mixed-effects logistic regression further reveals significant effects of register and semantic group, alongside substantial variation across individual *N₂* lemmas, with lemma-level variation making the strongest contribution to constructional choice. These findings demonstrate that the two competing evaluative constructions are functionally related but not fully interchangeable and that their competition reflects the interplay of lexical, semantic, and contextual factors.

Keywords: Evaluative Binominal Construction; evaluative morphology; constructional competition; register variation; multifactorial study

1. Introduction

Subjective evaluation can be expressed through different constructional patterns. In English, evaluative meanings can be conveyed through the phrasal Evaluative Binominal Construction (EBC) and the evaluative compound construction. The EBC (cf. ten Wolde 2023; Van Goethem 2023; Van Goethem et al. forthc.) is commonly realized through the constructional scheme [*DET₁ N₁ PREP DET₂ N₂*], as in *a hell of a job* and *a gem of a book*. In this pattern, *N₁* typically contributes the evaluative meaning and functions as an evaluative modifier, while *N₂* denotes the entity being evaluated. Evaluation is thus attributed to the referent denoted by *N₂* within an expanded nominal structure. In evaluative compounds [*N₁ N₂*], by contrast, the first element contributes an evaluative or intensifying meaning to the head noun (Štekauer 2015),

as in *hell job* and *nightmare scenario*. In this pattern, evaluation forms part of the internal conceptual characterization of the nominal unit, with N_1 contributing to the description or classification of the entity denoted by N_2 . The two constructions therefore package evaluation differently, while both may serve to express intensified evaluation involving a noun.

The relationship between such functionally related alternatives has often been described in terms of competition. According to Berg (2014: 342), competition may be defined as “the struggle among alternative forms during the retrieval process”. In language use, competition arises when alternative forms are available for selection because of their similar suitability for a particular goal. In a related sense, De Smet et al. (2018: 198) describe functionally similar expressions as forms that “can be used more or less interchangeably in the same discourse contexts because their semantic extensions cover overlapping areas of conceptual space.” Importantly, competing expressions need not have identical meanings; rather, their meanings need only be sufficiently similar to allow interchangeability in practice (De Smet et al. 2018: 206).

Following this understanding, competition does not presuppose complete semantic equivalence or unrestricted interchangeability. Rather, it concerns cases where formally distinct expressions overlap in function while their distribution may be shaped by linguistic and contextual factors. The relationship between the EBC and the evaluative compound construction is examined in this sense as a case of competition. Although the two constructions differ in how evaluation is packaged, their shared evaluative function places them within an overlapping domain.

Among the nouns occurring in the N_1 position of evaluative constructions, *hell* provides a particularly informative case for examining such competition. Ten Wolde’s (2023: 110) corpus-based analysis identifies *hell* as the most frequent N_1 both synchronically and diachronically in the dataset and as the most productive N_1 in evaluative constructions in Present-Day English. These findings establish *hell* as a central element of evaluative constructions and motivate closer examination of its use across the two patterns.

At the same time, the evaluative contribution of *hell* varies across contexts. The evaluative use of *hell* is related to, but distinct from, its literal meaning. In its literal sense, *hell* refers to “the infernal regions regarded in various religions as a place of suffering and evil” and is strongly associated with Christian traditions of punishment and suffering.

These negative and affective associations provide a semantic basis for its non-literal extensions. The *Oxford English Dictionary*, for instance, records uses of *hell* expressing “annoyance, anger, or surprise” (*Oxford English Dictionary* n.d.). Accordingly, *hell* may function as a referential noun in literal uses, form part of a formulaic expression such as *go to hell*, occur in the WHX construction, such as *what the hell*, which has been widely studied from various linguistic perspectives (Hugou 2017; Ippolito 2024; Kim & Kim 2023), or contribute to evaluative meaning in expressions such as *a hell of a problem* (cf. ten Wolde 2023).

The compound pattern introduces an additional source of variation, since the evaluative interpretation of [*HELL*(-)N₂] depends not only on *hell* but also on the noun with which it combines. Some combinations permit a transparent evaluative interpretation, whereas others express literal (non-evaluative) and often conventionalized meanings, as illustrated by dictionary-attested expressions such as *hell-fire*, *hell money*, and *hell mouth* (*Oxford English Dictionary* n.d.). The degree of functional overlap between [*a HELL of a/an* N₂] and [*HELL*(-)N₂] therefore cannot be assumed from formal similarity alone but needs to be established from their actual patterns of use.

Previous research on evaluative constructions has identified differences between phrasal and compound patterns in productivity as well as in lexical and semantic preferences (cf. Van Goethem 2023, 2026). However, the extent to which these differences are conditioned by contextual factors, particularly register, remains less explored. This concern is consistent with usage-based variationist research, which treats alternation as multifactorial and examines how multiple linguistic and contextual factors jointly condition language users’ choices between functionally related alternatives through quantitative methods (e.g., Wang et al. 2024). Among these conditioning factors, register is particularly relevant as an important language-external factor. Recent corpus-based variationist research has demonstrated its role by identifying “a significant main effect of register under multivariate control or substantial random effects” (Szmrecsanyi 2019: 4).

Corpus-based register studies further show that linguistic features vary systematically across registers at lexical, grammatical, and lexico-grammatical levels (e.g., Biber 2012). Since registers differ in situational characteristics such as participants, participant relationships, mode, processing circumstances, setting, communicative purposes, and topic (Biber

& Conrad 2019), such contextual variation may influence the distribution of competing constructions. Accordingly, the choice between [*a HELL of a/an* N₂] and [*HELL(-)*N₂] may be shaped not only by lexical and semantic factors, but also by the contexts in which the constructions occur.

The present study therefore investigates the competition between the phrasal and compound constructions by examining their distribution and the factors conditioning constructional choice, with particular attention to register, individual N₂ lemmas, and broader semantic patterns among N₂ nouns. To this end, Section 2 reviews previous work on evaluative constructions and constructional competition; Section 3 describes the corpora and methods; Section 4 presents the quantitative results; Section 5 discusses the findings; and Section 6 outlines the study's conclusions and limitations.

2. Literature Review

2.1 Evaluative Binominal Construction

In Construction Grammar, Goldberg (2006: 5) defines a construction as follows: “any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency.”

The EBC can be situated within the broader family of evaluative *of*-binominal constructions discussed in previous research (e.g., Kim & Sells 2015; ten Wolde 2023). As described by Van Goethem et al. (forthc.), it follows the schematic pattern [DET₁ N₁ PREP DET₂ N₂], in which the prepositional slot is fixed while the other slots allow variation. Expressions such as *that hell of a day*, *a dream of a day*, and *a hell of a job* can therefore be regarded as instances of this construction. Semantically, the two nouns are interconnected through evaluative semantics: N₁ expresses “an extraordinary (physical or abstract) quality, often in a metaphorical way”, while N₂ represents “the entity to which this quality is assigned and, as such, functions as the referential head of the construction” (translated from Verhagen 2005: 202). The construction as a whole thus displays a non-compositional, subjective-evaluative meaning. Related patterns have been discussed under various labels, including “evaluative binominal noun phrase”,

“binominal noun phrase”, “expressive binominal noun phrase”, “qualitative binominal noun phrase”, and have been documented across a range of languages (ten Wolde 2023: 64).

Within this broader research context, *hell* is a particularly important case in the development of English *of*-binominal constructions. Ten Wolde (2023: 106–110) traces the grammaticalization and subjectification of *hell* across a family of six related *of*-binominal constructions. In this development, *hell* shifts from a lexical first noun evoking “a biblical hell” to increasingly evaluative and intensifying functions. In the evaluative binominal noun phrase, it denotes negative properties, as in *a hell of a calling*; in the later evaluative modifier construction, it expresses positive evaluation, as in *a hell of a man*; and in the binominal intensifier construction, it functions primarily as an intensifier, as in *a hell of a good time*. Ten Wolde (2023: 113) characterizes this development as a shift in semantic prosody from negative to positive, accompanied by increasing subjectification, as *hell* comes to encode speaker-oriented subjective evaluation and ultimately develops into a degree intensifier.

Hugou’s (2017) synchronic corpus-based study of the *what the hell* (WHX) construction further demonstrates the role of *hell* in expressing emotion and intensity. Drawing on Huddleston and Pullum’s description of *the hell* as an “emotive modifier”, Hugou (2017: 6–7) claims that its presence “merely enhances the strength of the illocutionary force of the speech act” and allows speakers to “overtly express their stance”. Based on COCA data, Hugou (2017: 25) also shows that the WHX construction is highly productive, with a range of syntactic and lexical variants (e.g., *wh-the-hell*, *wh-in-the-hell*, *wh-in-hell*, *bloody hell*, and *heck*). These variants differ in their perceived intensity and pragmatic force. Although Hugou’s study does not examine the EBC directly, it shows that the WHX construction displays a semantic preference for religion-related words and that its evaluative and intensifying meanings are shaped by the interaction between cultural, contextual, and constructional factors.

Van Goethem et al.’s (forthc.) study is more directly relevant because it investigates the EBC through a corpus-based, cross-linguistic constructionist analysis of Dutch, English, French, and Italian. Using the *TenTen* corpus family, they identified the 15 most frequent EBC micro-constructions per language. For each micro-construction, a CQL query specifying the N1 lemma was used to retrieve a random sample of 1000 concordance lines, which were

manually checked until 100 relevant instances had been retained. This yielded a standardized dataset of 1,500 relevant occurrences per language. The dataset was annotated for formal properties, evaluative meaning (*ameliorative*, *pejorative*, *augmentative*, and *diminutive*), the semantic category of N_2 (*person*, *animal*, *object*, *place*, and *concept*), and N_2 lemma. The latter formed the basis for the calculation of productivity scores proposed by Baayen (2009), namely Realized Productivity, i.e. a measure of lexical diversity in the N_2 -slot, and Hapax Legomena-Token Ratio or Potential Productivity (PP), i.e. a measure of the expandability of the N_2 -slot (Van Goethem et al. *forthc.*).

In the English data from *enTenTen21*, *hell* is the most frequent N_1 and expresses a mix of pejorative and ameliorative values (Van Goethem et al. *forthc.*). English EBCs also show an overall preference for N_2 nouns denoting abstract concepts (Van Goethem et al. *forthc.*). More broadly, this study identifies an inverse correlation between realized productivity and semantic coherence. Realized productivity is measured by Type-Token ratio, calculated from the N_2 lemmas occurring in the standardized 100-token sample for each micro-construction. Semantic coherence, following Barðdal's (2008: 27) notion of "the internal consistency between the members of the schema", is operationalized as "the percentage of N_2 lemmas belonging to the largest semantic category within each pattern" (Van Goethem et al. *forthc.*).

The English EBCs are distributed along a productivity-coherence continuum. At one end, *a beast of a N_2* is highly productive but semantically less coherent, occurring with N_2 from all five semantic categories. At the other end, *a stale of a N_2* is less productive but more coherent, with 97% of the N_2 referring to *concept* (Van Goethem et al. *forthc.*). *A hell of a N_2* occupies a position between these two extremes, combining a TTR of 0.60 with a semantic coherence score of 0.80 (Van Goethem et al. *forthc.*).

2.2. Evaluative Compounding and Constructional Competition

Compounds and multi-word expressions (MWE) are not categorized in the same way across languages. In languages where compounds and MWEs are treated as distinct word-formation types, Finkbeiner and Schlücker (2019: 1) raise the question of whether they "act in a relation of competition or complementation in the formation of new lexical units".

In English, however, compounds are often treated as one type of MWE, although the boundaries between compounds, other MWEs, and syntactic phrases are not always clear-cut (Bauer 2019; Finkbeiner & Schlücker 2019). Consequently, central questions in the analysis of English compounds and MWEs concern their formal distinction, including prosodic, morphological, and syntactic properties, as well as the relation between compounding and MWE formation (Finkbeiner & Schlücker 2019). Bauer's (2019: 55–56) discussion of [N+P+N] constructions illustrates this gradience: some are established MWEs, such as *sense of humour*, whereas others are more freely generated syntactic combinations, such as *piece of cheese*. This general issue is also relevant to evaluative language, where evaluation may be expressed through both compound and phrasal constructions. More generally, evaluative meaning can be encoded through various word-formation processes, such as compounding, affixation, reduplication, and sound symbolism (Štekauer 2015: 47–53).

The relationship between evaluative compound and phrasal constructions has been examined in Dutch by Van Goethem (2023), who compares nominal ameliorative compounds such as *een droomhuis* 'a dream house' with evaluative binominal phrases such as *een droom van een huis* 'a dream of a house'. The findings based on four concrete case studies show that the two patterns differ in structural flexibility and productivity. The phrasal construction allows more complex slot fillers and greater creativity, as illustrated by expressions such as *een schat van een Engelse nanny* 'a sweetheart of an English nanny' and novel expressions such as *een droom van een stofzuiger* 'a dream of a vacuum cleaner' and *een pracht van een ribeye steak* 'a beauty of a ribeye steak'. In contrast, the compound pattern is more structurally constrained (e.g., *een wonderfiets* 'a wonder bike'), more lexicalized, and, in some cases, semi-fixed, as shown by forms such as *een wonderkind* 'a prodigy child'.

The two constructions also differ in their semantic preferences. The phrasal pattern tends to favour animate referents, including people and animals (e.g., *een droom van een moeder* 'a dream of a mother'), while the compound pattern is more strongly associated with inanimate referents, including places and abstract concepts (e.g., *een prachtkans* 'a beautiful opportunity'). Taken together, Van Goethem (2026: 62) proposes that the two constructions exhibit "a functional division of labour".

The Dutch findings raise the question of whether a similar differentiation can be observed between evaluative compound and phrasal constructions in English. In this respect, *hell* is a particularly relevant case. Although *hell* has been widely discussed in *of*-binominal constructions, WHX constructions, and other expressive patterns, its use as the evaluative first element of [N+N] compounds has received comparatively little attention. Consequently, whether the phrasal construction [*a HELL of a(n) N₂*] and the compound construction [*HELL(-)N₂*] differ systematically in their distribution, productivity, and lexical and semantic preferences remains underexplored, particularly across registers. To address this gap, the present study compares the two constructions in terms of their distribution, productivity, and lexical and semantic preferences across registers, and further examines how lexical, semantic, and register-related factors condition constructional choice. The study therefore addresses the following research questions:

- RQ1: How do the two competing evaluative constructions vary in distribution and productivity across registers when they occur with the same head nouns?
- RQ2: Which N₂ lemmas and semantic domains are significantly associated with each construction?
- RQ3: To what extent do individual N₂ lemmas, semantic groups, and register influence the choice between the phrasal and compound constructions?

3. Method

This study adopts a corpus-based constructionist approach to investigate the competition between two English evaluative constructions containing *hell*. To answer the three research questions, concordance lines of the two constructions were retrieved from the corpora described in Section 3.1 and organized into one English competition dataset. Its compilation and analytical procedures are presented in Sections 3.2 and 3.3.

3.1 Corpus Data

The English competition dataset was compiled from two corpora available through Sketch Engine (Kilgarrieff et al. 2014): the English Web 2021 corpus (*enTenTen21*), part of the *Ten-Ten* web-corpus family (Jakubiček et al. 2013; *Sketch Engine* n.d.), and the *OpenSubtitles 2018 Parallel Corpus* (Lison et al. 2018). Seven English sub-corpora representing distinct registers

were included: *reference/encyclopedia*, *blog*, *discussion*, *fiction*, *news*, *legal*, and *movie subtitles*. The six *enTenTen21* sub-corpora consist of web-based written texts, whereas the *OpenSubtitles2018* sub-corpus consists of translated movie subtitles representing mediated audiovisual dialogue. These sources differ in their compilation, provenance, text structure, sampling procedures, duplication possibilities, communicative status, and temporal composition. However, all sub-corpora are processed using the same part-of-speech tagset, namely the English TreeTagger PoS tagset with Sketch Engine modifications, ensuring consistency in corpus querying and linguistic analysis.

Movie subtitles were included because they provide access to interactional and speech-like communication, offering a relevant context for examining evaluative and emotionally expressive language. In addition, subtitles corpora have been widely used in cross-linguistic studies of swearwords and taboo-languages (e.g., Díaz-Pérez 2020; Jiang & Doherty 2026), which makes them particularly relevant for the present study. Therefore, the combination of written and speech-like communicative contexts enables the study to investigate constructional variation across different communicative contexts, from formal to informal and from informational to narrative contexts.

Table 1 summarizes the sub-corpora included in the study, together with their sizes and representative registers. All corpus data were accessed through Sketch Engine, while the data processing and quantitative analyses reported in Sections 3.2–3.3 were conducted in R version 4.4.1 using RStudio.

Tab. 1: Overview of (Sub)Corpora Used

Corpus	Size	Register
English Web 2021 (<i>enTenTen21</i>)–reference/encyclopedia	3,047,342,438	Reference/encyclopedia
English Web 2021 (<i>enTenTen21</i>)–blog	1,569,499,442	Blog
English Web 2021 (<i>enTenTen21</i>)–discussion	2,103,533,595	Discussion
English Web 2021 (<i>enTenTen21</i>)–fiction	1,030,493,934	Fiction
English Web 2021 (<i>enTenTen21</i>)–news	2,420,719,017	News
English Web 2021 (<i>enTenTen21</i>)–legal	652,370,863	Legal
OpenSubtitles 2018 Parallel–English sub-corpus	821,040,909	Movie subtitles

Note. The *OpenSubtitles 2018* Parallel–English sub-corpus was extracted using the criterion that the language spoken in the film is English.

3.2 Compilation of the English Competition Dataset

Table 2 presents the four Corpus Query Language (CQL) searches used to extract the target constructions from the corpora. For the phrasal construction, two forms were included: the full form [*a HELL of a/an N₂*] and the reduced form [*HELLUVA N₂*]. The latter was included and grouped with the phrasal construction in the present analysis because it represents a reduced lexicalization historically related to *hell of a* and shares its intensifying evaluative function (*Oxford English Dictionary* n.d.). For the compound construction, two orthographic variants were included: the spaced form [*HELL N₂*] and the hyphenated form [*HELL-N₂*]. The lemma *lot* was excluded from the N₂ slot in the CQL searches in order to avoid cases where *lot* forms part of the expression [*a lot of + N*], since in such cases *lot* is not the target N₂ head noun. Table 3 reports the raw frequencies returned by these four queries in the two corpora.

After extraction, the phrasal and compound datasets were cleaned separately in two stages. In the automatic cleaning stage, N₂ forms were lowercased, stripped of extra spaces, cleared of non-alphabetic characters, and then lemmatized. Shared N₂ lemmas were subsequently identified by comparing the cleaned phrasal and compound datasets. A lemma was treated as shared if it occurred in both construction types. This procedure yielded 601 shared N₂ lemmas.

The shared N₂ list was then manually examined to remove non-target items, unclear cases, and retrieval errors. First, the N₂ lemmas were manually examined, and concordance lines were consulted for unclear cases to determine whether their occurrences with *hell* represented instances of the target constructions. Second, dictionary entries were consulted for *hell*-initial compounds to identify established lexical items in which *hell* retained a literal or conventionalized meaning rather than an evaluative meaning (*Oxford English Dictionary* n.d.). These compounds were compared against the shared N₂ list, and non-target items (e.g., *hell box*, *hell-fire*, *hell god*, *hell money*, and *hell mouth*) were excluded. After this filtering stage, the final English competition dataset contained 510 shared lemmas and 16,186 rows (see Appendix I), which formed the basis for the analysis of English constructional competition. Table 4 summarizes the number of shared lemmas and dataset rows before and after manual filtering.

Tab. 2: Target Constructions and CQL Queries Used for Extraction

Construction	Target construction	Corpus Query Language
Full form	[Det ₁ <i>HELL</i> of Det ₂ N ₂]	[word="a an the"] [lemma="hell" & tag="NN"] [word="of"] [word="a an the"] [tag="NN NNS" & lemma!="lot"]
Reduced form	[Det <i>HELLUVA</i> N ₂]	[word="helluva"] [tag="NN NNS" & lemma!="lot"]
Spaced form	[<i>HELL</i> N ₂]	[lemma="hell" & tag="NN"] [tag="NN NNS"]
Hyphenated form	[<i>HELL</i> -N ₂]	[word="hell-[a-zA-Z]+" & tag="NN NNS"]

Tab. 3: Raw Frequencies of Target Constructions by Corpus

Construction	<i>English Web 2021</i>	<i>OpenSubtitles2018</i>	Total
Full form	8,474	8,109	16,583
Reduced form	2,542	217	2,759
<i>Subtotal</i>	11,016	8,326	18,166
Spaced form	10,000	5,632	15,632
Hyphenated form	5,852	813	6,665
<i>Subtotal</i>	15,852	6,445	22,297

Note. The original number of results for the spaced form of the compound construction in *English Web 2021* was 24,722. Because of retrieval limits, only 10,000 rows were extracted.

Tab. 4: Shared N₂ Lemmas and Competition Dataset Before and After Manual Filtering

Stage and Measure	Count
<i>After automatic cleaning</i>	
Rows in the two datasets	39,640
<i>Before manually filtering</i>	
Shared N ₂ lemmas	601
Rows in competition dataset	24,315
<i>After manual filtering</i>	
Shared N ₂ lemmas	510
Rows in competition dataset	16186

3.3 Analysis of English Constructional Competition

To answer RQ1, the English competition dataset was analysed by register distribution and construction type. For each register, raw frequencies of the phrasal and compound

constructions were counted and normalized per million words. Percentages were calculated to show the relative shares of the two constructions within each register. As a supplementary measure, log ratios were used to make constructional choice more directly comparable. A chi-square test was used to test whether the distribution of the two constructions differed significantly across registers, and Cramér's V was calculated to measure the strength of the association.

Token frequency and type frequency were first calculated for each construction, with token frequency referring to the total number of construction occurrences and type frequency referring to the number of distinct N₂ lemmas filling the constructional slot. Constructional productivity was then assessed following Baayen's (2009) framework. Three productivity measures were calculated, namely, Realized productivity, Expanding productivity, and Potential productivity. Realized productivity was calculated as the proportion of distinct N₂ lemma types relative to the total number of construction tokens (i.e., type–token ratio, TTR). Expanding productivity was calculated as the proportion of hapax N₂ lemmas (i.e., N₂ lemmas occurring only once within a given construction in a given register) relative to the total number of hapax types in the corresponding corpus sample. Potential productivity was operationalized as the proportion of hapax N₂ lemmas relative to the total number of construction tokens. These measures have been applied in usage-based approaches to constructional productivity (e.g., Zeldes 2012).

To answer RQ2, Distinctive Collexeme Analysis (DCA) was used to identify N₂ lemmas that showed a preference for one construction over the other. Only lemmas with a total frequency of at least five were included. DCA is a type of collostructional analysis that measures the association between lexical items (i.e., collexemes) and constructions (Gries & Stefanowitsch 2004). It is particularly suitable for comparing two functionally similar constructions because it identifies which collexemes are more strongly attracted to one construction than the other. This approach has been widely applied in corpus-based constructional studies (e.g., Pavlović 2020). The analysis was conducted with the `collex.dist()` function from the `collostructions` package.

The collexeme results were then combined with semantic annotations of the N₂ lemmas. The annotation process was assisted by the USAS online English tagger (<https://ucrelapi.lancaster.ac.uk/usas/tagger.html>), developed at Lancaster University

(<http://ucrel.lancs.ac.uk/wmatrix/>), which implements the UCREL Semantic Analysis System (USAS; Rayson et al. 2004). Its original tagset was based on McArthur's (1981) *Longman Lexicon of Contemporary English*, while the revised tagset is hierarchically organized into 21 major discourse fields, as illustrated in Figure 1. These major discourse fields are further divided into 232 category labels, as listed on the website (<https://ucrel.lancs.ac.uk/usas/>). For example, the discourse field T: Time includes four semantic subcategories, such as T1: Time, T2: Time: Beginning and ending, T3: Time: Old, new and young; age, and T4: Time: Early/late. The subcategory T1: Time can be further subdivided into six more specific labels, including T1.1: Time: General, T1.2: Time: Momentary, and T1.3: Time: Period.

In the USAS output, each item is assigned to one syntactic tag and one or more semantic tags. When more than one tag is possible, the tags are presented in approximate order of likelihood, with the most probable tag appearing first. For the present analysis, the first USAS tag assigned to each collocate was used as the basis for semantic annotation, as this represents the most probable interpretation in the USAS output. This principle was also applied when alternative tags were provided, such as Q2.1/S2, in which case only the first tag (Q) was retained. The retained USAS tags were then mapped onto broader semantic domains, which were used to classify the N₂ lemmas. Semantic domains were counted according to constructional preference, and their proportions were calculated within each construction. A chi-square test and Cramér's V were likewise employed to assess the association between semantic domain and constructional preference. The standardized residuals were subsequently examined to identify domains that were over- or under-represented in each construction.

A general and abstract terms	B the body and the individual	C arts and crafts	E emotion
F food and farming	G government and public	H architecture, housing and the home	I money and commerce in industry
K entertainment, sports and games	L life and living things	M movement, location, travel and transport	N numbers and measurement
O substances, materials, objects and equipment	P education	Q language and communication	S social actions, states and processes
T Time	W world and environment	X psychological actions, states and processes	Y science and technology
Z names and grammar			

Fig. 1: USAS Semantic Tagset

To answer RQ3, mixed-effects binary logistic regression modelling was used to assess the effects of individual head nouns, semantic groups, and register on the alternation between the phrasal and compound constructions. Construction type served as the binary dependent variable, with N_2 lemma entered as the only random effect and modelled as a random intercept. Semantic domain and register were included as fixed effects. The reference levels were selected in light of the findings for RQ1 and RQ2: *World and environment* for semantic domain and *fiction* for register.

To facilitate statistical modelling and interpretation, the 21 USAS semantic-domain levels were subsequently conflated into four broader semantic groups. This procedure was motivated by two considerations. First, several USAS domains contained relatively few observations, which could lead to unstable estimates when modelled as separate predictors. Second, a preliminary additive mixed-effects model was fitted with the original 21 semantic domains. The estimated coefficients from this model were examined together with the observed constructional distributions of individual domains. Domains showing similar directions of association with constructional choice and similar constructional distributions were grouped together. This grouping was intended to reduce the number of predictor levels while preserving meaningful semantic distinctions, rather than to optimize model performance.

The resulting four groups, i.e. (1) *entity- and environment-oriented domains*, (2) *institutional, material, and process domains*, (3) *human, social, and evaluative domains*, and (4) *abstract*,

temporal, communicative, and activity domains, provided broader semantic categories for regression analysis. The final labels reflect the dominant semantic characteristics shared by the included USAS domains. In addition, the *entity- and environment-oriented domains* group, which contained the original reference level *World and environment*, was retained as the semantic reference category. Table 5 presents the grouping scheme, while Appendix III provides the full list of USAS domains assigned to each group.

In order to achieve the optimal model, the modelling process began with a null model containing only a random intercept for N₂ lemma. A main-effects model with register and semantic group as predictors was then fitted, followed by an interaction model incorporating semantic group $\times$ register interaction. The models were compared using likelihood-ratio tests and Akaike Information Criterion (AIC), with lower AIC values indicating better fit. The performance of the final model was evaluated using marginal and conditional R² values, calculated with the *MuMIn* package, together with the in-sample area under the receiver operating characteristic curve (AUC). The final model included semantic group, register, their interaction, and a random intercept for N₂ lemma. All models were fitted in R using the *lme4* package, following a modelling procedure adapted from Wang et al. (2024).

Tab. 5: Semantic grouping scheme for the mixed-effects logistic regression analysis

Semantic group	No. of labels	Estimate range	Total N	Phrasal %	Tendency
Entity- and environment-oriented domains	4	< 0	1,071	23.7%	compound-oriented
Institutional, material, and process domains	4	0.2–0.6	2,709	31.2%	slightly compound-oriented
Human, social, and evaluative domains	6	0.6–1.0	3,135	76.9%	slightly phrasal-oriented
Abstract, temporal, communicative, and activity domains	7	> 1.0	9,271	81.6%	phrasal-oriented

Note. The second column reports the number of semantic labels included in each semantic group. For the *entity- and environment-oriented domains*, the estimate range includes the reference level and negative estimates. Phrasal % refers to the proportion of phrasal constructions within each semantic group.

4. Results

4.1 Distribution and Productivity

As shown in Table 6, the two constructions show clear differences in both overall frequency and relative preference across registers. Their combined normalized frequency is highest in *movie subtitles*, followed by *fiction* and *discussion*, but substantially lower in the *reference/encyclopedia* and *legal* registers. Despite this variation in overall frequency, the phrasal construction predominates in all registers. Within the phrasal construction, [*a HELL of a/an* N₂] accounts for the majority of tokens in most registers, while the reduced variant [*HELLUVA* N₂] represents a smaller proportion of the dataset and varies across registers. Its contribution is more substantial in registers such as *discussion* and *blog*, where it represents 9.78% and 13.57% of all constructional tokens, respectively, compared with only 1.53% in *movie subtitles*.

Strong preferences for the phrasal construction occur in *movie subtitles* (log ratio = 1.657) and *discussion* (log ratio = 1.162), both of which are associated with spoken or interactional discourse. *Reference/encyclopedia* also shows a clear phrasal preference (log ratio = 2.053), although the two constructions are relatively infrequent in this register overall. By contrast, *fiction* displays a relatively weak preference for the phrasal construction (log ratio = 0.279), while *blog*, *news*, and *legal* registers show more moderate phrasal preference. The marked predominance of the phrasal construction in *movie subtitles* is noteworthy, given the greater formal brevity of the compound pattern. This finding is discussed further in Section 5.1.

These descriptive patterns are supported by the chi-square test, which confirms that the relative use of the phrasal and compound constructions varies significantly across registers ($\chi^2 = 452.14$, $p < .001$). However, the Cramér's *V* value of 0.167 suggests that the association between construction type and register is small (Cohen 1988). The small association indicates that register contributes to the alternation between the two constructions but does not fully account for constructional choice.

Tab. 6: Normalized Frequencies and Relative Distribution of the Two Constructions by Register

Register	Total pmw	Phrasal %		Compound %	Log ratio
		Full	Reduced		
Movie subtitles	7.710	74.39	1.53	24.08	1.657
Fiction	2.519	47.61	7.2	45.18	0.279
Discussion	1.464	59.34	9.78	30.89	1.162
Blog	1.371	50.23	13.57	36.20	0.817
News	0.694	51.70	11.61	36.69	0.787
Reference/encyclopedia	0.107	69.02	11.66	19.33	2.053
Legal	0.037	54.17	12.5	33.33	0.957

Note. pmw = per million words. Full and reduced refer to the [*a HELL of a/an* N₂] and the reduced variant [*HELLUVA* N₂], respectively. Positive log-ratio values indicate a higher relative proportion of the phrasal construction, whereas negative values indicate a higher relative frequency of the compound construction.

Table 7 presents the type and hapax type distributions underlying the productivity analysis, while Figure 2 illustrates the productivity profiles of the two constructions across registers. Both constructions show relatively low realized and potential productivity across most registers, although considerably higher values are observed in *legal* texts and, to a lesser extent, in *reference/encyclopedia*. Their productivity profiles nevertheless differ across registers and measures, with the compound construction generally exhibiting higher realized and potential productivity than the phrasal construction. This general tendency is observed across registers, although register-specific differences are apparent. *Movie subtitles* display relatively low realized and potential productivity for both constructions despite their high frequency of occurrence, whereas *fiction* represents an exception in that the phrasal construction shows slightly higher expanding productivity, while the two constructions have identical potential productivity scores (0.074). Expanding productivity more generally shows the opposite tendency, with the phrasal construction exceeding the compound construction in most registers. *Legal* texts show considerably higher realized and potential productivity scores, which should be interpreted cautiously due to the small number of observations. The representative N₂ lemmas in Table 8 further illustrate differences in the lexical fillers associated with the two constructions. Across registers, the compound construction tends to

occur with a wider range of N₂ lemmas, whereas the phrasal construction tends to cluster around frequent nouns such as *time* and *job*.

Taken together, these findings show that the two evaluative constructions differ in both register distribution and productivity. Although the phrasal construction is generally more frequent across registers, the compound construction tends to exhibit higher realized and potential productivity. This contrast suggests that the two patterns may also differ in their preferred N₂ lemmas and the semantic domains associated with these fillers, which are examined in the following section.

Tab. 7: N₂ lemma types and hapax types by register

Register	Phrasal Types	Phrasal Hapax Types	Compound Types	Compound Hapax Types
Movie subtitles	294	65	144	46
Blog	249	131	198	128
Discussion	261	116	237	130
Fiction	220	106	208	87
Legal	11	7	6	5
News	182	105	162	105
Reference/encyclopedia	92	53	26	16

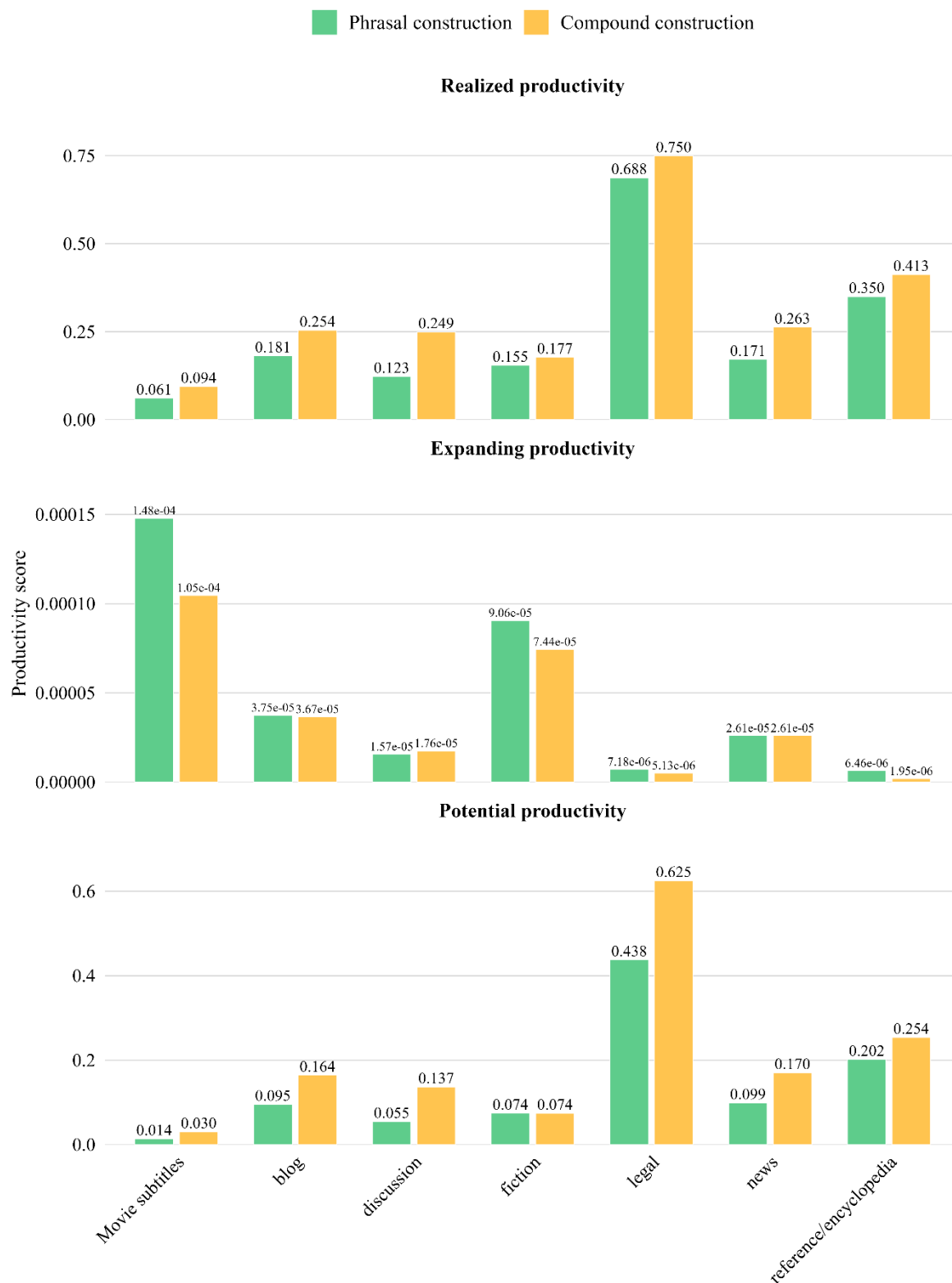


Fig. 2: Constructional productivity across registers

Tab. 8: Representative N₂ Lemmas by Register and Construction

Register	Construction	Representative N ₂ lemmas
Discussion	Compound	<i>hole</i> (211), <i>week</i> (108), <i>right</i> (31), <i>man</i> (30), <i>break</i> (29)
	Phrasal	<i>time</i> (700), <i>job</i> (169), <i>deal</i> (120), <i>difference</i> (70), <i>thing</i> (67)
Fiction	Compound	<i>hole</i> (334), <i>beast</i> (71), <i>cat</i> (50), <i>nightmare</i> (35), <i>right</i> (33)
	Phrasal	<i>time</i> (316), <i>job</i> (113), <i>fight</i> (111), <i>day</i> (72), <i>thing</i> (64)
News	Compound	<i>hole</i> (245), <i>week</i> (34), <i>break</i> (22), <i>house</i> (15), <i>match</i> (13)
	Phrasal	<i>time</i> (44), <i>job</i> (21), <i>ride</i> (14), <i>thing</i> (12), <i>game</i> (8), <i>guy</i> (6)

Note. Values in parentheses indicate the frequency of each N₂ lemma within the corresponding register and construction type.

4.2 Lexical and Semantic Preferences

This section turns from register-level distribution to lexical-semantic preference by identifying the N₂ lemmas significantly associated with each construction and comparing their distribution across the predefined USAS semantic domains. Of the 336 N₂ lemmas occurring at least five times in both constructions, DCA identified 158 with a significant constructional preference. As shown in Table 9, 78 were associated with the phrasal construction and 80 with the compound construction. Although the compound pattern was associated with slightly more significant collexemes overall, both constructions displayed distinctive lexical preferences across all significance levels.

Tab. 9: Distribution of Identified Collexemes by Construction Type and Significance Level

Construction	Significance level					Total number of collexemes
	*	**	***	****	*****	
Phrasal construction	31	11	8	3	25	78
Compound construction	16	13	11	10	30	80
Total	47	24	19	13	55	158

Note. Asterisks indicate significance levels in the collexeme analysis. The significant collexemes are grouped into five levels according to the Fisher's exact test *p*-values obtained from the DCA: * *p* < .05, ** *p* < .01, *** *p* < .001, **** *p* < .0001, and ***** *p* < .00001.

Tab. 10: Top 20 distinctive collexemes for each construction

Phrasal construction		Compound construction	
Collexeme	COLL.STR.LOGL	Collexeme	COLL.STR.LOGL
time	1108.33833	hole	3453.03743
job	786.99123	week	426.67611
fight	271.79117	right	335.64391
thing	182.66601	break	273.27166
day	169.41233	house	240.95159
ride	134.09747	beast	226.29473
story	130.89683	home	189.54803
mess	123.60133	cat	156.89794
deal	121.58549	bitch	123.38616
party	117.38492	sound	81.41544
show	104.05982	man	55.68587
game	95.01896	world	54.81202
guy	74.85276	talk	52.3556
drug	64.35127	dude	52.3556
year	57.56757	business	50.87093
team	52.07538	nightmare	49.32964
place	38.86657	difference	46.15804
idea	37.52931	try	44.66965
cop	30.47595	trial	44.01515
season	29.70401	diver	39.51407

Note. COLL.STR.LOGL refers to collostruction strength on the negative log-transformed scale.

Table 10 provides a closer view of these lexical preferences by listing the 20 most distinctive collexemes for each construction. *Time* and *job* show the strongest association with the phrasal pattern, while *hole* and *week* rank highest for the compound pattern. The

compound pattern also attracts several nouns with salient evaluative or interpersonal associations, such as *bitch*, *dude*, and *nightmare*, whereas the phrasal pattern favours relatively general nouns, including *time*, *job*, and *thing*. These preliminary lexical contrasts are examined in greater detail through the semantic-domain analysis that follows.

Figure 3 extends the lexical comparison by showing how the distinctive collexemes associated with each construction are distributed across the USAS semantic domains. The phrasal construction is represented in 15 of the 21 domains, whereas the compound construction occurs in 19, indicating broader semantic coverage for the latter. The phrasal profile is also more concentrated. Four domains—*General and abstract terms*, *Social actions, states and processes*, *Language and communication*, and *Movement, location, travel and transport*—account for 51.2% of its distinctive collexemes. In comparison, the compound construction shows a more dispersed distribution, with its top five domains – *Social actions, states and processes*, *Psychological actions, states and processes*, *General and abstract terms*, *Life and living things*, and *Substances, materials, objects and equipment* – jointly accounting for 48.7%.

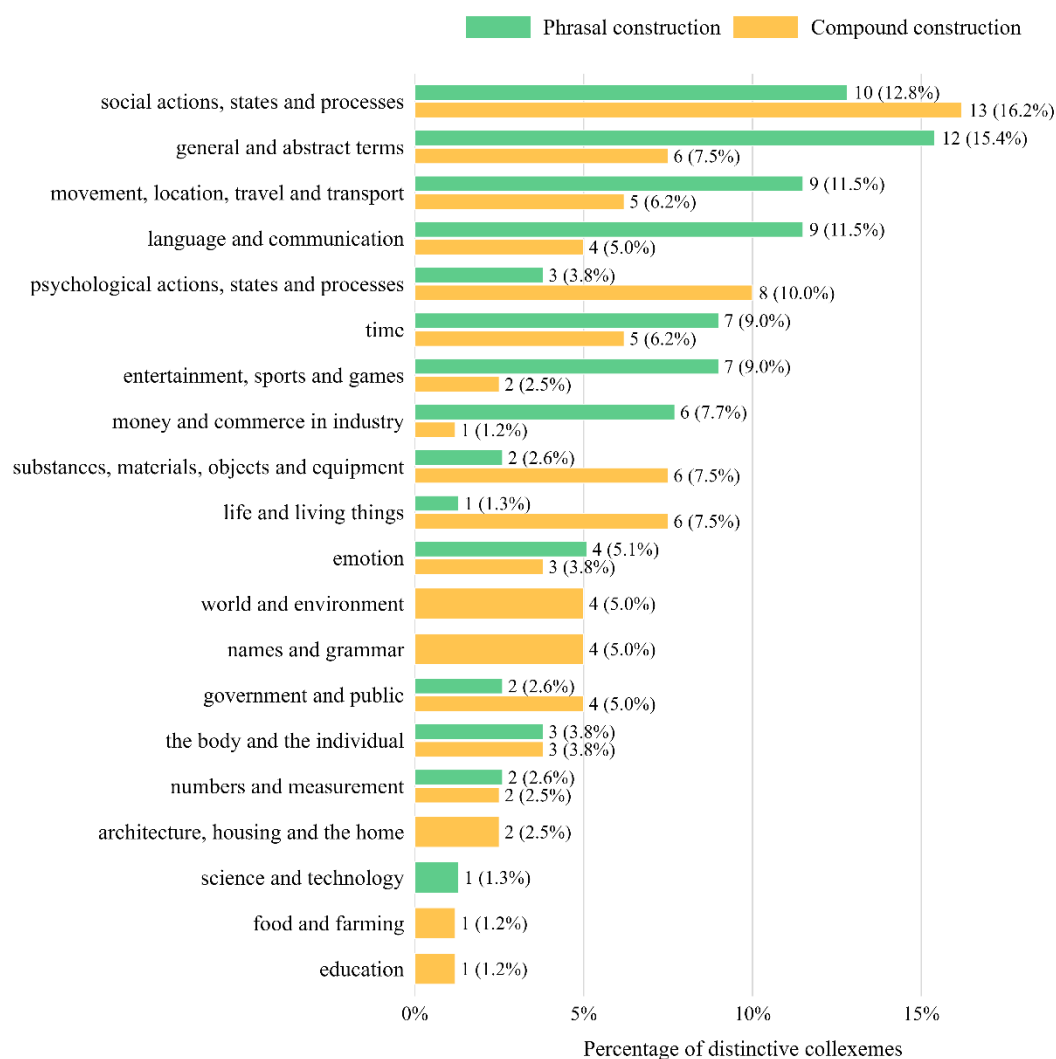


Fig. 3: Distribution of Distinctive Collexemes across USAS Semantic Domains by Construction

The two constructions differ in the relative prominence of semantic domains. The phrasal construction has higher proportions in eight domains, with the largest difference found in *General and abstract terms* (+7.9 percentage points). The compound construction, by contrast, is more strongly represented in five other semantic domains, particularly *Psychological actions, states and processes* and *Life and living things*, each exceeding the phrasal construction by 6.2 percentage points.

Some semantic domains are represented exclusively among the distinctive collexemes associated with one construction. Five semantic domains are represented only among the compound-attracted collexemes: *World and environment* (e.g., *world*, *planet*, *wind*, and *rain*), *Names and grammar* (e.g., *right* and *something*), *Architecture, housing and home* (e.g., *house* and *home*), *Food and farming* (e.g., *drink*), and *Education* (e.g., *train*). By

contrast, *Science and technology* (e.g., *engineer*) is represented solely among the phrasal-attracted collexemes.

These descriptive patterns point to differences in the semantic profiles of the two constructions. The compound construction is distributed across a broader range of domains, while the phrasal construction is concentrated in a smaller set. To determine whether this distributional contrast was statistically significant, a chi-square test of independence was conducted on the 158 collexeme types. The results revealed a significant association between semantic domain and construction preference ($\chi^2 = 33.774$, $p < 0.05$), with a Cramér's V value of 0.462 indicating a relatively large association (Cohen 1988). Standardized residuals were then examined to identify the semantic domains that contributed most to the significant association. The complete DCA results and supplementary semantic domain statistics are provided in Appendix II.

As shown in Table 11, only *Names and grammar* and *World and environment* Life reached the standardized-residual threshold of 2, indicating their overrepresentation among collexemes associated with the compound construction. *Names and grammar* includes nouns such as *right*, *fest*, *something*, and *slayer*. No domain reached this threshold for the phrasal construction, although the strongest tendencies were found in *Money and commerce in industry* (e.g., *reputation*), *Entertainment, sports and games* (e.g., *party*), *General and abstract terms* (e.g., *chance*), *Language and communication* (e.g., *speech*), and *Movement, location, travel and transport* (e.g., *truck*).

Tab. 11: Semantic Domains Most Strongly Associated with the Compound and Phrasal Constructions

Construction	Semantic domain	Standardized residual
<i>Compound construction</i>	Names and grammar	2.0003
	World and environment	2.0003
	Life and living things	1.8990
	Psychological actions, states and processes	1.5195
	Substances, materials, objects and equipment	1.4148
<i>Phrasal construction</i>	Money and commerce in industry	1.9675
	Entertainment, sports and games	1.7555
	General and abstract terms	1.5596
	Language and communication	1.4953
	Movement, location, travel and transport	1.1695

Note. Positive standardized residuals indicate that a semantic domain is more strongly represented than expected in the corresponding construction; values greater than 2 indicate notable over-representation.

Together, Sections 4.1 and 4.2 have established register-related, lexical, and semantic differences between the two constructions. Section 4.3 therefore evaluates their respective contributions to constructional choice through mixed-effects logistic modelling.

4.3 Lexical, Semantic, and Register Effects on Constructional Choice

The mixed-effects logistic regression model was fitted to the English competition dataset, with construction type as the binary outcome. The compound construction was set as the reference category, so the model estimated the likelihood of the phrasal construction being selected. Positive estimated coefficients represent higher odds of the phrasal construction, whereas negative coefficients represent lower odds. Exponentiated coefficients are reported as odds ratios (OR), which serve as measures of effect size. Values above 1 indicate increased odds of the phrasal construction relative to the relevant reference level, while values below 1 indicate decreased odds. Interaction effects are interpreted using model-predicted probabilities.

As summarized in Table 12, the main-effects model provided a significantly better fit than the null model ($\chi^2(9) = 230.82, p < .001$). Adding the interaction between semantic group and register produced a further significant improvement in model fit ($\chi^2(17) = 108.13, p < .001$). The interaction model also yielded the lowest AIC value of 10306.45, although BIC favoured the more parsimonious main-effects model. Given the significant likelihood-ratio test and the theoretical relevance of the interaction to RQ3, the interaction model was selected as the final model.

Tab. 12: Comparison of Mixed-Effects Logistic Regression Models

Model	AIC	BIC	logLik	Deviance	χ^2	Δdf	p value
Null model	10593.40	10608.78	-5294.70	9482.56	—	—	—
Main-effects model	10380.58	10465.19	-5179.29	9334.19	230.82	9	< .001 ***
Interaction model	10306.45	10521.82	-5125.22	9230.25	108.13	17	< .001 ***

Note. BIC = Bayesian information criterion; logLik = log-likelihood. The χ^2 , Δdf , and p values are from likelihood-ratio tests comparing each model with the preceding model. Significance codes: *** $p < .001$.

The model evaluation statistics for the final interaction model are reported in Table 13. The marginal R^2 value was 0.085, representing the variance explained by the fixed effects of semantic group, register, and their interaction. The conditional R^2 value was 0.402, representing the variance explained by the full model, including both the fixed effects and the random intercept for N_2 lemma. The substantial difference between the two values suggests that the fixed effects account for only a modest proportion of the variation in constructional choice, which is unsurprising given the small number of fixed-effects predictors included in the model and the substantial lexical variation across N_2 lemmas. The higher conditional R^2 indicates substantial lemma-level variation in shaping the alternation between the two constructions. This interpretation is further supported by a supplementary diagnostic comparison, which found that adding a random intercept for N_2 lemma reduced model deviance by 53.04% relative to an intercept-only logistic regression without random effects. The model also achieved an AUC of 0.936, indicating strong discrimination between the two constructions in the present dataset.

Tab. 13: Model evaluation statistics for the final interaction model

Model	Marginal R^2	Conditional R^2	In-sample AUC
Final interaction model	.085	.402	.936

Adapted from the original statistical results, Table 14 presents significant fixed effects from the final interaction model. The phrasal construction is the outcome category being modelled. The significant negative intercept shows that the reference condition—*entity- and environment-oriented domains in fiction*—favours the compound construction, as illustrated in (1).

- (1) Reference condition (*entity- and environment-oriented domains in fiction*)
Sometimes they're good and sometimes they're little **hell beasts**.

Tab. 14: Statistically significant fixed effects in the final interaction model

Predictor	Estimate	SE	OR	z value	p value
Intercept	-1.275	0.294	0.28	-4.335	< .001 ***
<i>Main effect: Semantic group</i>					
Institutional, material, and process domains	1.180	0.372	3.25	3.169	.002 **
Human, social, and evaluative domains	1.136	0.340	3.12	3.343	< .001 ***
Abstract, temporal, communicative, and activity domains	2.123	0.326	8.36	6.511	< .001 ***
<i>Main effect: Register</i>					
Discussion	1.051	0.328	2.86	3.199	.001 **
Movie subtitles	1.023	0.266	2.78	3.843	< .001 ***
News	0.890	0.384	2.43	2.316	.021 *
<i>Interaction effect</i>					
Institutional, material, and process domains × Discussion	-1.119	0.415	0.33	-2.694	.007 **
Human, social, and evaluative domains × Discussion	-0.959	0.368	0.38	-2.605	.009 **
Abstract, temporal, communicative, and activity domains × Discussion	-0.862	0.352	0.42	-2.446	.014 *
Human, social, and evaluative domains × Movie subtitles	0.761	0.309	2.14	2.464	.014 *
Abstract, temporal, communicative, and activity domains × Movie subtitles	-0.608	0.290	0.54	-2.096	.036 *
Institutional, material, and process domains × News	-1.302	0.497	0.27	-2.620	.009 **
Abstract, temporal, communicative, and activity domains × News	-0.918	0.411	0.40	-2.232	.026 *

Note. SE = standard error; OR = odds ratio. Significance codes: *** $p < .001$; ** $p < .01$; * $p < .05$.

The main effects of semantic group show that, in *fiction*, all three non-reference semantic groups have significantly higher odds of occurring in the phrasal construction than *entity- and environment-oriented domains*. The strongest effect is found for *abstract, temporal, communicative, and activity domains*, with an odds ratio of 8.36. This means that, in *fiction*, the odds of the phrasal construction are 8.36 times higher for *abstract, temporal communicative, and activity domains* than for the reference category *entity- and*

environment-oriented domains. *Human, social, and evaluative domains* and *institutional, material, and process domains* also show higher odds of the phrasal construction, with odds ratios of 3.12 and 3.25, respectively. These semantic contrasts are illustrated in (2).

(2) Semantic main effects in *fiction*

- a. When we stormed the Initiative, we had **a hell of a time** getting in. (*abstract, temporal, communicative, and activity domains*)
- b. Becca is my neighbor, a cop and **a hell of a woman**. (*human, social, and evaluative domains*)
- c. I have to clean **a hell of a mess** out of my car. (*institutional, material, and process domains*)

Register also contributes significantly to constructional choice. Within the reference category of *entity- and environment-oriented domains*, three registers reach statistical significance: *discussion* (OR = 2.86), *movie subtitles* (OR = 2.78), and *news* (OR = 2.43). Compared with *fiction*, these registers show higher odds of the phrasal construction. This suggests that, within *entity- and environment-oriented domains*, *fiction* is relatively more compound-oriented, whereas *discussion*, *movie subtitles*, and *news* are more phrasal-oriented, as shown in (3).

(3) Register main effects for *entity- and environment-oriented domains*

- a. Turns out this guy lived **a hell of a life**. (*discussion*)
- b. There's **a hell of a storm** coming up. (*movie subtitles*)
- c. Punk said this was **a hell of a mountain** to climb. (*news*)

The significant interactions show that the effect of semantic group on constructional choice varies across registers. Relative to the reference combination of *entity- and environment-oriented domains* in *fiction*, one positive interaction is found for *human, social, and evaluative domains* in *movie subtitles*. In this register, this semantic group is associated with a stronger tendency towards the phrasal construction than would be expected from the corresponding main effects alone, as illustrated in (4).

- (4) Positive interactions: *human, social, and evaluative domains* × *movie subtitles*
That's **a hell of a team** you've built.

The remaining significant interactions are negative. These involve *abstract, temporal, communicative, and activity domains* in *movie subtitles, discussion, and news*; *institutional, material, and process domains* in *discussion and news*; and *human, social, and evaluative domains* in *discussion*. These negative coefficients indicate that the phrasal tendency associated with the relevant semantic groups is weaker in these registers than would be expected from the corresponding main effects alone. This probabilistic interpretation is illustrated in (5), where both constructions occur within the same semantic-group–register combination.

(5) Negative interactions

Abstract, temporal, communicative, and activity domains in movie subtitles

- a. What is this, **hell week**?
- b. Oh, I had **a hell of a job** finding this place.

Institutional, material, and process domains in news

- c. The **hell cannon** could be moved there to support of the militant advance on the depot.
- d. It may only be episode one of a much larger story, but it delivers **a hell of a package** in spite of it.

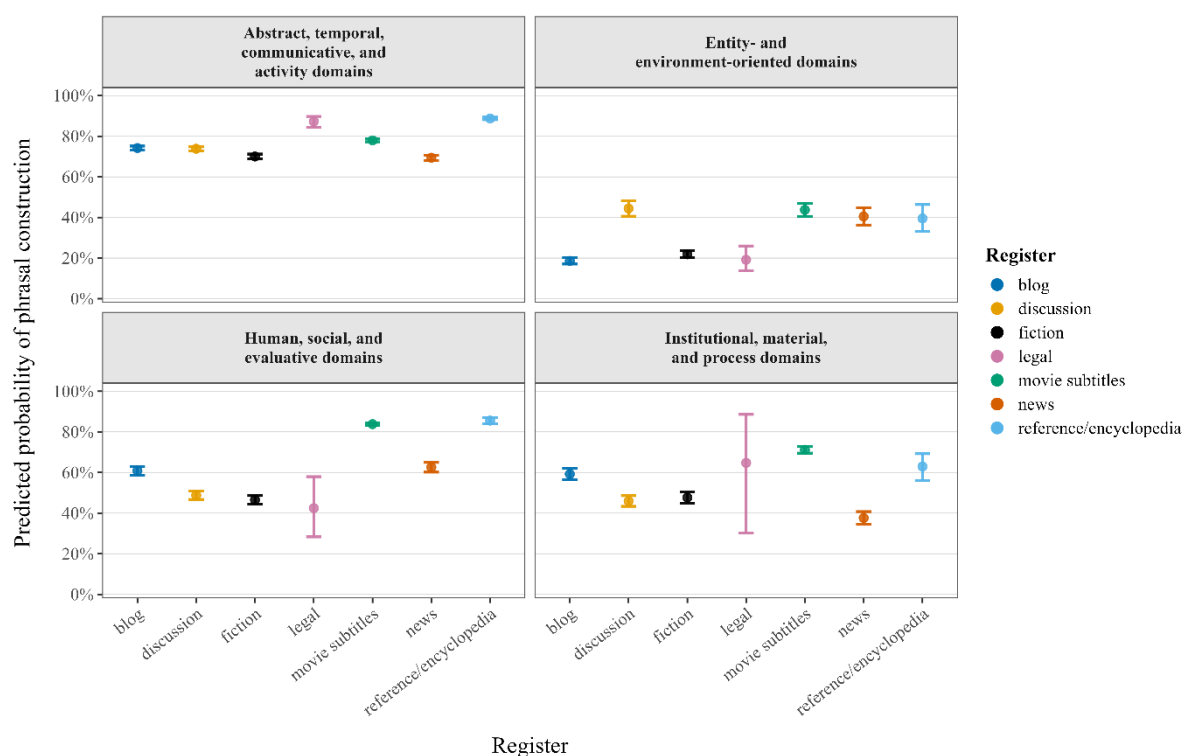


Fig. 4: Predicted probability of the phrasal construction by semantic group and register

To clarify the interaction patterns, Figure 4 presents the model-predicted probabilities of the phrasal construction for each semantic group across registers. *Abstract, temporal, communicative, and activity domains* show consistently high predicted probabilities of the phrasal construction across all registers, with the highest estimate in *reference/encyclopedia* and *legal*, as illustrated in (6a). The estimate for *legal* texts should be interpreted cautiously because this register contains relatively few observations.

- (6) Model-predicted interaction patterns
- a. It was **a hell of a ride** but as they say all good things must end.
(*abstract, temporal, communicative, and activity domains* in *reference/encyclopedia*)
 - b. When Ortiz leaves tonight, he can hold his head high. He gave the fans **a hell of a fight**.
(*human, social, and evaluative domains* in *reference/encyclopedia*)
 - c. It's **a hell of a plan**, isn't it?
(*institutional, material, and process domains* in *movie subtitles*)
 - d. Neither am I, I just want to help this gentleman and then I'm going to walk out of this **hell hole** with my head held high.
(*institutional, material, and process domains* in *movie subtitles*)

By contrast, *entity- and environment-oriented domains* generally have lower predicted probabilities, reflecting a stronger tendency towards the compound construction. *Human, social, and evaluative domains* are more sensitive to register, with comparatively high probabilities in *movie subtitles* and *reference/encyclopedia*, as shown in (6b), but lower probabilities in *fiction* and *discussion*. A similar degree of register variation is found for *institutional, material, and process domains*, which show the strongest phrasal tendency in *movie subtitles*, as in (6c), although compound realizations are also attested as in (6d).

As an interim summary, the regression analysis reveals that N_2 lemmas, semantic groups, and register all contribute to constructional choice, with additional variation arising from the interaction between semantic group and register. The following Discussion section considers the role of register in constructional competition, the lexical and semantic differentiation between the two constructions, and the relative contributions of these factors to constructional choice.

5. Discussion

5.1 Register Variation in Constructional Competition

The findings for RQ1 show that the phrasal and compound constructions differ across registers in both frequency and productivity. Both constructions are most frequent in *movie subtitles* and *fiction*, but occur far less often in *reference/encyclopedia* and *legal* registers. This distribution can be related to differences in the situational characteristics and communicative purposes of these registers: *movie subtitles* represent spoken-like and interaction-oriented discourse, while *fiction* provides narrative contexts in which evaluation may contribute to affective description and characterization. By comparison, *reference/encyclopedia* and *legal* texts are more formal or institutionalized and may therefore provide fewer contexts for strongly evaluative expressions. Despite these differences in overall frequency, the phrasal construction predominates across all registers, although the strength of this preference varies considerably. The relative preference is strongest in *reference/encyclopedia* and *movie subtitles*, whereas *fiction* shows only a weak phrasal preference, with the two constructions more evenly distributed. This variation may also partly reflect differences in the distribution of reduced variants within the phrasal construction.

Productivity, however, follows a different pattern from frequency. Although the phrasal construction is more frequent across registers, the compound construction generally exhibits higher productivity scores, particularly in terms of hapax-based productivity measures. This finding differs from Van Goethem's (2026) observations for Dutch, where phrasal evaluative patterns are reported to be more productive and creative, whereas compounds tend to be more constrained and lexicalized or semi-fixed. In the present data, the compound construction shows greater productivity despite its lower overall frequency, suggesting that its N₂ slot remains relatively open to lexical expansion.

This relationship between frequency and productivity varies across registers, with *fiction* and *movie subtitles* illustrating two different patterns. In *fiction*, the two constructions display relatively similar productivity profiles, with the phrasal construction showing slightly higher expanding productivity. This relatively balanced productivity profile is consistent with the distribution of lexical fillers observed in this register, where both constructions occur with a

range of N_2 lemmas, including compound forms such as *hell game/ship/break* and *hell-hole*, alongside phrasal expressions such as *hell of + job/scene/place* and *helluva card/smile/team*. This pattern may reflect the nature of fictional discourse, where both dialogue and narration provide contexts for diverse evaluative expressions. By contrast, *movie subtitles* show a different relationship between frequency and productivity. This register has the highest overall frequency of the two constructions and strongly favours the phrasal construction (75.92%). Given that movie language, like face-to-face conversation, is “involved, non-narrative and situation-dependent” (Forchini 2019: 145), and that subtitles are subject to spatial and temporal constraints, the compact compound construction might be expected to play a stronger role because it allows evaluative meaning to be packaged into a shorter nominal form. The results, however, show that brevity alone does not determine the constructional choice. Despite the pressure for economy, the phrasal construction remains dominant. This suggests that evaluative expressions in *movie subtitles* are shaped not only by formal constraints but also by their interactional functions in dialogic contexts. In particular, phrasal forms (e.g., *hell of + thing/voice/dress/puzzle* and *helluva difference/time/job/place*) may provide a more explicit pattern for expressing speaker stance, whereas compound forms (e.g., *hell dream/business/ship/example* and *hell-hole/fight*), although more compact and lexically flexible, occur less frequently in this register.

Overall, register is associated with differences in both the frequency and productivity of the two constructions, but these dimensions do not vary in parallel. The contrasting patterns in *fiction* and *movie subtitles* further demonstrate that greater frequency does not necessarily correspond to greater realized and potential productivity. The generally low productivity values also indicate that much of the competition is concentrated around a restricted set of recurrent N_2 lemmas. These register-level patterns therefore provide only part of the picture, as constructional choice may also depend on the lexical and semantic properties of the N_2 slot, which are examined in the following section.

5.2 Lexical and Semantic Differentiation between the Competing Constructions

The DCA results reveal that the two constructions differ systematically in the N_2 lemmas they attract. Of the 336 N_2 lemmas meeting the frequency threshold for DCA, 158 display a significant preference for one pattern. Thus, the two constructions are not freely

interchangeable even when they occur with the same head noun. This supports Gries and Stefanowitsch's (2004: 102) view that distinctive collexemes are expected to match proposed constructional semantics and that construction-specific preferences are partly semantically motivated.

These lexical preferences form broader semantic patterns. Although the two constructions attract a similar number of significant collexemes, with 80 associated with the compound construction and 78 with the phrasal construction, these collexemes are not distributed in the same way across semantic domains. The phrasal collexemes occur in fewer USAS domains and are concentrated in several domains. However, no individual domain is significantly overrepresented for the phrasal construction. By contrast, the compound construction spans more semantic domains and shows significant overrepresentation in *Names and grammar* and *World and environment*. The compound construction therefore displays broader semantic coverage while also showing particularly strong association with specific domains, whereas the phrasal construction shows narrower coverage without significant overrepresentation in any individual domain.

Such a pattern may be related to the different ways in which the two constructions package evaluation. In the phrasal construction, evaluation is imposed on the referent denoted by N₂, which may allow the construction to combine with nouns from different semantic areas without developing strong associations with particular domains. By contrast, in the compound construction, evaluative meaning contributes to the conceptualisation of the nominal unit, which may facilitate greater lexical-semantic diversification. Overall, the two constructions are systematically differentiated in their lexical and semantic preferences, suggesting that they occupy overlapping but non-equivalent regions of evaluative meaning, with constructional competition partly conditioned by the properties of the N₂ slot.

5.3 Relative Contributions of Lexical, Semantic, and Register Factors

The regression analysis provides a more integrated view of how lexical, semantic, and contextual factors contribute to the alternation between the two constructions. The large difference between the marginal R² and conditional R² values indicates that lemma-level variation makes a substantial contribution to the explained variation. This points to strong

lexical conditioning, as individual N₂ lemmas vary substantially in their preference for the phrasal or compound construction.

Notably, semantic group continues to contribute to constructional choice when register and lemma-level variation are taken into account. *Entity- and environment-oriented domains* favour the compound construction, consistent with the RQ2 finding that *Names and grammar* and *World and environment* were overrepresented among compound-attracted collexemes. In contrast, *abstract, temporal, communicative, and activity domains* favour the phrasal construction, in line with Van Goethem et al.'s (forthc.) observation that English EBCs preferentially occur with nouns denoting abstract concepts. This pattern partly parallels the functional division of labour reported for Dutch evaluative constructions, where phrasal patterns favour animate referents and compounds favour inanimate ones. The English compound construction shows a comparable preference for entity- and environment-oriented meanings, whereas the English phrasal construction differs by favouring abstract, temporal, communicative, and activity-related meanings rather than animate referents.

Register also contributes to constructional choice, although its effect is not uniform across semantic groups. Within the reference semantic group, *discussion*, *movie subtitles*, and *news* favour the phrasal construction relative to *fiction*. This pattern suggests that the phrasal pattern may be more compatible with contexts in which evaluation is closely tied to interaction, expressivity, and stance-taking. The compound construction, by contrast, appears better suited to more compact nominal description. This interpretation is broadly consistent with Dušková's (2009: 29) finding that pre-modifiers in fiction more often serve descriptive or evaluative functions than those in academic prose. The interaction effects further suggest that semantic preference are register-dependent. In *movie subtitles*, human, social, and evaluative meanings appear particularly compatible with the phrasal construction, which may reflect the dialogic and affective nature of this register. In *discussion* and *news*, by contrast, the weaker phrasal tendencies observed for several semantic groups suggest that constructional choice is shaped by the interaction between meaning and context.

Taken together, these findings highlight the dynamic nature of evaluative language, with constructional preferences varying across lexical, semantic, and communicative contexts.

6. Conclusion

This study explored the competition between the phrasal construction [*a HELL of a/an* N₂] and the compound construction [*HELL* N₂] across registers. Although the two patterns share an evaluative function, they differ systematically in their distribution, productivity, and lexical and semantic preferences. Among the factors examined, lemma-level variation makes the strongest contribution to constructional choice, while semantic group and register also exert significant effects. The interaction further indicates that semantic preferences vary across registers. Their alternation is therefore best understood as constrained competition rather than free variation.

Several limitations should be considered when interpreting these findings. First, although manual cleaning was conducted, the English competition dataset may still contain extraction errors or irrelevant instances. In addition, retrieval restrictions in the English Web 2021 corpus limited the number of available results for the spaced form of the compound construction, resulting in an incomplete dataset. The relatively small number of observations in the legal register also limits the reliability of conclusions drawn for that category. The semantic classification of N₂ lemmas, assisted by the USAS online tagger, may contain errors, particularly for polysemous or context-dependent items.

Moreover, the broader semantic groups used in the regression model were partly derived from the observed relationships between semantic domains and constructional choice in the corpus, which may have introduced a degree of circularity. Future research could apply theoretically motivated semantic classifications and compare the results with those obtained here. Finally, the comparison between registers is potentially affected by corpus or source differences. Specifically, the *movie subtitles* register is represented by *OpenSubtitles2018*, whereas the other registers are drawn from *enTenTen21*. Consequently, differences between *movie subtitles* and other registers may partly reflect corpus/source effects rather than register differences alone.

These limitations notwithstanding, the study provides a more nuanced understanding of how evaluative meaning is distributed across competing constructions and communicative contexts.

AI Statement

This paper has made use of ChatGPT (version 5.5) for language editing and stylistic refinement. All AI-assisted outputs have been critically reviewed and revised by the authors to ensure accuracy and academic integrity.

References

- Baayen, Rolf Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics: An international handbook*, 900–919. Berlin & New York: Mouton de Gruyter.
- Barðdal, Jóhanna. 2008. *Productivity: Evidence from Case and Argument Structure in Icelandic*. Amsterdam & Philadelphia: John Benjamins. DOI: [10.1075/cal.8](https://doi.org/10.1075/cal.8).
- Bauer, Laurie. 2019. Compounds and multi-word expressions in English. In Barbara Schlücker (ed.), *Complex Lexical Units: Compounds and Multi-Word Expressions*, 45–68. De Gruyter. DOI: [10.1515/9783110632446-002](https://doi.org/10.1515/9783110632446-002).
- Berg, Thomas. 2014. Competition as a unifying concept for the study of language. *The Mental Lexicon* 9(2). 338–370. DOI: [10.1075/ml.9.2.08ber](https://doi.org/10.1075/ml.9.2.08ber).
- Biber, Douglas. 2012. Register as a predictor of linguistic variation. *Corpus Linguistics and Linguistic Theory* 8(1). 9–37.
- Biber, Douglas & Susan Conrad. 2019. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd edition. New York: Routledge. DOI: [10.4324/9780203771587](https://doi.org/10.4324/9780203771587).
- De Smet, Hendrik, Frauke D’hoedt, Lauren Fonteyn & Kristel Van Goethem. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29(2). 197–234. DOI: [10.1515/cog-2016-0025](https://doi.org/10.1515/cog-2016-0025).
- Díaz-Pérez, Francisco Javier. 2020. Translating swear words from English into Galician in film subtitles: A corpus-based study. *Babel* 66(3). 393–419. DOI: [10.1075/babel.00162.dia](https://doi.org/10.1075/babel.00162.dia).
- Dušková, Libuše. 2009. Noun modification in fiction and academic prose. *Brno Studies in English* 35(2). 29–51.
- Finkbeiner, Rita & Barbara Schlücker. 2019. Compounds and multi-word expressions in the languages of Europe. In Barbara Schlücker (ed.), *Complex Lexical Units: Compounds and Multi-Word Expressions*, 1–44. Berlin & Boston: De Gruyter. DOI: [10.1515/9783110632446-001](https://doi.org/10.1515/9783110632446-001).
- Forchini, Pierfranca. 2019. Dimensions “Assembled”: The Nature of Movie Conversation. *ITA*. <https://hdl.handle.net/10807/146267> (accessed 28 September 2026).

- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Gries, Stefan Thomas & Anatol Stefanowitsch. 2004. Extending collocation analysis: A corpus-based perspective on ‘alternations’. *International Journal of Corpus Linguistics* 9(1). 97–129. DOI: [10.1075/ijcl.9.1.06gri](https://doi.org/10.1075/ijcl.9.1.06gri).
- Hugou, Vincent. 2017. The WHX construction (what the hell...?) and intensity. A corpus-based study. *Lexis. Journal in English Lexicology* (10). DOI: [10.4000/lexis.1103](https://doi.org/10.4000/lexis.1103).
- Ippolito, Michela. 2024. The Hell With Questions. *Journal of Semantics* 41(1). 53–76. DOI: [10.1093/jos/ffae001](https://doi.org/10.1093/jos/ffae001).
- Jakubiček, Miloš, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý & Vít Suchomel. 2013. The TenTen Corpus Family. <https://is.muni.cz/publication/1120431/en/The-TenTen-Corpus-Family/Jakubiček-Kilgarriff-Kovar-Rychly> (accessed 28 September 2026).
- Jiang, Qihang & Doherty, Stephen. 2026. A corpus-based study on the English-Chinese swearwords produced by amateur (fansubbing) and professional (prosubbing) subtitling. *Perspectives* 34(3). 447–473. DOI: [10.1080/0907676X.2024.2328755](https://doi.org/10.1080/0907676X.2024.2328755).
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1). 7–36. DOI: [10.1007/s40607-014-0009-9](https://doi.org/10.1007/s40607-014-0009-9).
- Kim, Jungsoo & Jong-Bok Kim. 2023. Aggressively non-D-linked construction and ellipsis: A Direct Interpretation approach. *Journal of Linguistics* 59(2). 257–291. DOI: [10.1017/S0022226722000226](https://doi.org/10.1017/S0022226722000226).
- Kim, Jong-Bok & Peter Sells. 2015. English binominal NPs: A construction-based perspective. *Journal of Linguistics* 51(1). 41–73. DOI: [10.1017/S002222671400005X](https://doi.org/10.1017/S002222671400005X).
- Lison, Pierre, Jörg Tiedemann & Milen Kouylekov. 2018. *OpenSubtitles2018: Language Resources and Evaluation Conference. Proceedings of the 11th International Conference on Language Resources and Evaluation* (LREC 2018), 1742–1748.
- Oxford English Dictionary. n.d. *Helluva*. DOI: [10.1093/OED/2911562528](https://doi.org/10.1093/OED/2911562528).
- Oxford University Press. n.d. *Hell*, n. & int. DOI: [10.1093/OED/4718197636](https://doi.org/10.1093/OED/4718197636).
- Pavlović, Vladan. 2020. *Too early to say: The English too ADJ to V construction and models of cross-cultural communications styles: A collocation approach*. *International Journal of Corpus Linguistics* 25(3). 297–321. DOI: [10.1075/ijcl.19044.pav](https://doi.org/10.1075/ijcl.19044.pav).
- Rayson, Paul, Dawn Archer, Scott Piao & Tony McEnery. 2004. The UCREL Semantic Analysis System. *Proceedings of the Workshop on Beyond Named Entity Recognition Semantic Labelling for NLP Tasks*, in Association with LREC-04, 7–12.
- Sketch Engine. n.d. enTenTen—English corpus from the web | Sketch Engine. Sketch Engine. https://www.sketchengine.eu/ententen-english-corpus/?utm_source=chatgpt.com (accessed 22 May 2026).
- Štekauer, Pavol. 2015. Word-Formation Processes in Evaluative Morphology. In Nicola Grandi & Livia Körtvélyessy (eds.), *Edinburgh Handbook of Evaluative Morphology*, 43–60. Edinburgh: Edinburgh University Press.
- Szmrecsanyi, Benedikt. 2019. Register in variationist linguistics. *Register Studies* 1(1). 76–99. DOI: [10.1075/rs.18006.szm](https://doi.org/10.1075/rs.18006.szm).
- ten Wolde, Elnora. 2023. *The English Binominal Noun Phrase: A Cognitive-Functional Approach*. Cambridge: Cambridge University Press. DOI: [10.1017/9781108921893](https://doi.org/10.1017/9781108921893).

- Van Goethem, Kristel. 2023. *Een pracht van een constructie vs. een prachtconstructie* [A beauty of a construction vs. a beauty construction]. *Nederlandse Taalkunde* 28(3). 325–349. DOI: [10.5117/NEDTAA2023.3.004.GOET](https://doi.org/10.5117/NEDTAA2023.3.004.GOET).
- Van Goethem, Kristel. 2026. Evaluative compounding in Dutch: Productivity and competition in simile-based constructions. *SKASE Journal of Theoretical Linguistics* 23(4). 49–65.
- Van Goethem, Kristel, Valentina Piunno & Isa Hendrikx. forthc. How do productivity, semantic coherence and creativity interconnect? Cross-linguistic evidence from the Evaluative Binominal Construction. *Creativity at the intersection of Construction Grammar and Phraseology (Constructional Approaches to Language)*. Amsterdam & Philadelphia: John Benjamins.
- Wang, Menglan, Guiying Jiang & Yan Cheng. 2024. A Corpus-Based Multifactorial Study of *Help/help to* Alternation in Learners' Language: From the Perspective of Probabilistic Grammar. *Sage Open* 14(4). DOI: [10.1177/21582440241293535](https://doi.org/10.1177/21582440241293535).
- Zeldes, Amir. 2012. *Productivity in Argument Selection. From Morphology to Syntax*. Berlin & Boston: De Gruyter Mouton. DOI: [10.1515/9783110303919](https://doi.org/10.1515/9783110303919).

Lin Ding

Université catholique de Louvain

Place Cardinal Mercier 31

1348 Louvain-la-Neuve, Belgium

lin.ding@student.uclouvain.be

Kristel Van Goethem

F.R.S.-FNRS & Université catholique de Louvain

Institut Langage et Communication

Place Cardinal Mercier 31

1348 Louvain-la-Neuve, Belgium

kristel.vangoethem@uclouvain.be

Ferran Suñer Muñoz

Université catholique de Louvain

Institut Langage et Communication

Place Cardinal Mercier 31

1348 Louvain-la-Neuve, Belgium

ferran.sunermunoz@uclouvain.be



This is an open access publication. This work is licensed under a Creative Commons Attribution CC-BY 4.0 license. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>