Toward a corpus-based identification of nominal relationality and uniqueness: A constructionist approach

Kazuho KAMBARA
Ritsumeikan University / Shiga, Japan
kazy0324@pep-rg.jp

Taishi CHIKA
Kyoto University / Kyoto, Japan
nanou7614@gmail.com

Abstract

A distribution-based classification of nouns has attracted many scholars (de Bruin and Scha, 1988; Barker, 2011). In forming a nominal phrase, some nouns (e.g., relational noun, kinship terms) take an "argument" using various kinds of grammatical construction (e.g., *the mother of Alice*, *Alice's head*). In contrast, some nouns can be realised with or without such elements (e.g., *Alice*, *a dog*). Löbner proposed a rigorous classification system by employing two criteria: (i) uniqueness of a referent, and (ii) relationality of a referent. The former corresponds to whether a referent can be identified as a unique entity (e.g., *Alice*), while the latter to whether a referent must be identified in relation to another entity (e.g., *mother*). Inspired by Löbner (2011), Glass (2022) demonstrated a quantitative approach to relationlity. However, the methodology taken in the conventional study had two limitations: (i) direction of attractions (i.e., attraction from the word to construction, and from the construction to the word), and (ii) interaction of relationality and uniqueness. We conducted an exploratory quantitative analysis based on collostructional approach (Gries and Stefanowitsch, 2004a,b; Stefanowitsch and Gries, 2003, 2005; Gries, 2019, 2022c,b). As a result, we succeeded in identifying typical instances of relational and non-relational nouns while distinguishing low-frequency items.

1 Introduction

In classifying nominal expressions, distributional patterns of expressions are often used. One of the most common distributional patterns is **relationality**. For instance, when the noun *mother* functions as a head of a nominal phrase, it usually takes a preposition phrase (or possessive phrase) as a "complement" (e.g., *the mother of Alice*, *Alice's mother*). These expressions are said to be relational if they tend to collocate with possessive expressions (de Bruin and Scha, 1988; Barker, 2011). Typical members of these classes of nouns include kinship terms (e.g., *brother, mother*) or body parts (e.g., *leg, head*). Löbner (2011) proposed a classification system which incorporates the relationality and **uniqueness** of referents. Glass (2022) demonstrated the effectiveness of a corpus-based approach to relationality.

Although Glass's results are promising, her analysis has two limitations: (i) the coverage of analysis and (ii) the need for accounting for the different directions of attraction (i.e., attraction from the word to the construction and the construction to the word). The former limitation corresponds to the limited types of grammatical constructions and nouns in consideration and the latter to the methodology of analysis. For instance, a typical relational noun *mother* can be used with or without the collocating possessive phrase in some contexts (e.g., *the mother*), suggesting that these nouns do not always require complements in every context. We show that the complex interaction of uniqueness and relationality can be observed by employing a collostructional analysis.

This paper reports the results of an exploratory quantitative analysis of the interactions of these possessive and non-possessive constructions. Section 2 overviews the nominal semantic classification system of Löbner (2011), and a quantitative approach to relational nouns (Glass, 2022) to motivate the current study. Section 3 presents the methods and procedure taken in this study. Section 4 presents the results and their interpretations to argue for the consistency of our results with previous studies. Section 5 summarises the paper.

2 Related studies

Section 2.1 briefly introduces Löbner's semantic class of nominal expressions and Glass's corpus-based evaluation of relationality. Section 2.2 points out possible improvements in evaluating the uniqueness and relationality of a given noun.

2.1 Concept types and determination

2.1.1 Four types of nominal expressions

In classifying nominal expressions, linguists use various grammatical constructions. For instance, the proper noun *Alice* is likely to be used without any articles (e.g., *? the Alice*), the noun *dog* with an article (e.g., *a dog*), and the noun *mother* with a possessive phrase (e.g., *Alice's mother*). Löbner (2011) proposed a rigorous system to classify nominal concepts in relation to the determination for referential use NP (i.e., a process of making up "full" NP). The values of the following features define the four classes of nouns.

(1) a. **Unique (abbreviated as [± U])**: [+U] iff a noun in question has the unique referent, [−U] otherwise.
b. **Relational (abbreviated as [± R])**: [+R] iff the referent of a noun must be given in relation to other entity (or, entities), [−R] otherwise.

For instance, the noun *stone* is assigned [−U] since the noun *stone* itself has no single referent, meaning that an article (e.g., *a*, *the*) is required to determine its referent in language use. These sortal nouns do not hold any inherent relations in their semantic contents. In contrast, relational nouns (e.g., kinship term [e.g., *mother*], body-parts [e.g., *hand*]), which hold inherent relation to other terms, are assigned [+R]. As shown in Table 1, these values yield four possible combinations: (i) SORTAL (i.e., [−U] & [−R]), (ii) INDIVIDUAL (i.e., [+U] & [−R]), (iii) RELATIONAL (i.e., [−U] & [+R]), and (iv) FUNCTIONAL (i.e., [+U] & [+R])[1].

Löbner lists syntactic environments in which a particular type of noun typically occurs (congruent determination) and might not occur in the usual context (incongruent determination), as shown in Table 2. It captures that any nominal can occur even in an incongruent determination by applying various type-shift rules.

2.1.2 Corpus based approach to relationality

Some studies (Löbner, 2011; Asmuth and Gentner, 2005) have attempted to identify relational nouns using substitution tests (i.e., syntactic tests) to see if the noun in question can occur in specific constructions (e.g., [Y'S X], [X OF Y]). Glass pointed out that the conventional method has difficulty in distinguishing sortal nouns (e.g., *dog, pencil*) from (clearly) relational ones (e.g., *sister, head*) since both of these types of nouns can occur in possessive constructions (e.g., *Alice's* {*dog, pencil, sister, head*}) due to the contextual factors.

Instead of conventional substitution tests, Glass argued for the effectiveness of corpus methods. Analysts can point out and address the continuous nature of the relationality of nouns by employing quantitative corpus methods. If a noun n_i in question is more likely to occur in the possessive construction than the other noun n_j, the relationality of n_i is higher than that of n_j. This conception was initially inspired by the work of Löbner (2011) and Nissim (2004).

The empirical survey results showed that kinship terms and body parts terms, considered typical members of relational nouns in literature, had the highest frequency of possessive constructions. In contrast, other categories (e.g., abstract concepts, natural objects) had lower frequencies. Based on those results, Glass argued for the validity of her hypothesis that *"nominal concept with greater human interaction have greater relationality"* [2].

2.2 Limitations of previous studies

Limitations of Glass's analysis are twofold: (i) the limited coverage of investigated constructions, and (ii) the necessity of refining collocational strength between words and construction types.

As discussed, Glass analyses only possessive constructions to account for the relationality of nouns, leaving out the uniqueness. Analysts (including linguists and researchers of NLP) can benefit from the classification system of Löbner (2011) if the coverage of analysis is sufficiently broad. Though Glass (2022) offers interesting empirical results, her analysis is limited. To refine her research, analysts should include constructions other than possessive constructions (e.g., *a X*, *the X*) to fully observe the effectiveness of Löbner's classification system.

To accurately account for the attraction/repelling relations between construction types and head nouns, analysts must consider two directions of

[1] Each type in Table 1 is defined as follows: SORTAL is defined as $\langle e, t\rangle$, INDIVIDUAL as e, RELATIONAL as $\langle e, \langle e, t\rangle\rangle$, and FUNCTIONAL as $\langle e, e\rangle$. For the reason of space, we do not discuss these processes in detail.

[2] In Glass's account, this hypothesis is supported by two auxiliary hypotheses. The first one is that *"nominal concepts with greater human interaction have higher token frequency"*. The other is that *"the higher the relationality, the more likely the noun is to be used in possessive constructions"*.

Table 1: Four types of nouns and their examples (Löbner, 2011, 307)

	Non-unique [−U]	Unique [+U]
Non-relational [−R]	SORTAL (e.g., *stone*, *book*, *adjective*, *water*)	INDIVIDUAL (e.g., *moon*, *weather*, *data*, *Maria*)
Relational [+R]	RELATIONAL (e.g., *sister*, *leg*, *part*, *attribute*)	FUNCTIONAL (e.g., *father*, *head*, *age*, *subject*)

Table 2: Congruent and incongruent mode of determination in four nominal types (Löbner, 2011, 307)

Values		Congruent	Incogruent
[−R]	[−U]	Indef., Plural, quantif., dem., absolute	singular definite, relational, possessive
[−R]	[+U]	singular definite, absolute	Indef., Plural, quantif., dem., relational, possessive
[+R]	[−U]	Indef., Plural, quantif., dem., relational, possessive	singular definite, absolute
[+R]	[+U]	singular definite, relational, possessive	Indef., Plural, quantif,., dem., absolute

attraction: the attraction from the construction to the word and the word to the construction (Gries, 2019, 392–393). **Collostructional analysis** can be employed to observe such measures. The term "collostructional analysis" refers to a family of collocational analyses that can accurately capture the collocational strength between grammatical constructions and words (Stefanowitsch and Gries, 2003, 2005; Gries and Stefanowitsch, 2004a,b; Gries, 2019, 2022c, 2023a,b).

Many association measures are proposed to measure the collocational strengths of a word and constructions (e.g., Dice, Jaccard, Mutual Information). However, these conventional measures come with limitations. Instead of computing a single value to capture the collocational strength[3], collostruction analysis uses $\Delta P_{cx \to w}$ and $\Delta P_{cx \to w}$ to capture the degrees of attraction (Gries, 2019, 2022c, 2023a). $\Delta P_{cx \to w}$ refers to the attraction strength from the construction to the word, and $\Delta P_{w \to cx}$ to the attraction strength from the word to the construction. Collostructional analysis subsumes a variety of collocation analysis (Hipert, 2014, 392). Of four well-known collostructional analyses, collexeme analysis should be effective in measuring the collocational strength between the head noun and each construction type (e.g., *My X, Y's X*).

In addition to the methodological advantages, employing collostructional analyses incorporates the assumptions of **construction grammar** (Hoffmann and Trousdale, 2013; Hoffmann, 2022), a theory that treats constructions, form-meaning pairings, as the primary unit of language knowledge. Since construction grammar subsumes various related theories, we do not commit to one specific sub-theory. However, our assumption is leaning towards a usage-based version of construction grammar (cf. Taylor, 2012)[4].

Suppose an analyst wishes to describe attraction relations between words and construction types to identify the semantic classes of nouns. In that case, she should analyse as many construction types as possible while accounting for the different degrees of attraction relations. In this study, we present a collostructional approach to uniqueness and relationality.

3 Methods

This section describes the employed method and procedure[5].

We extracted units of noun phrases to investigate how likely (or unlikely) a noun occurs in a given construction. For this purpose, we extracted units of noun phrases from the BNC component of the Treebank Semantic Parsed Corpus (TSPC) (Butler, 2022). We limited ourselves to using the BNC component because the frequency of NP construction is assumed to be high.

The total number of noun phrases yielded 13,288

[3] Computing multiple values to capture the attraction relations between words and constructions accurately is a recent approach (Gries, 2019, 2023b, 2022c). This approach is called **tupleziation** since it takes multiple measures into account (Gries, 2019, 394–396).

[4] Refer to Hoffmann and Trousdale (2013) for a finer-grained description of construction grammar.

[5] See data availability for the annotated data.

Table 3: Types of NP constructions

		HASART	
		TRUE	FALSE
HASPOSS	TRUE	Cx1	Cx2
	FALSE	Cx3	Cx4

cases. We excluded the following types in (2) to make the task manageable. As for (2-e), we annotated the parts of the whole noun phrase. For instance, if we encounter the case *"the salt and pepper"*, we separate the analysis for *"the salt"* and *"pepper"*. However, we included a complex noun with conjoined modifiers (e.g., *an expensive and boring book*)

(2)
- a. **Dates/Numbers/Length/Volume**: *Monday, Tuesday, ..., January, February, ..., 1, 2, ..., 1%, ..., 1cm, ...*
- b. **Pronouns**: *me, you, her, him, them, ..., everyone, anyone, ...*
- c. **ExtendedNouns**: *one, other, another, ..., former, less, ..., this, that, ..., the, ..., the best, ...*
- d. **Citation**: *Cruse (1986), ...*
- e. **ComplexNP**: *salt and pepper, ..., Alice and Bill, ..., Alice or Bill, ...*

These cleaning processes were conducted using regular expressions and manual annotations. The total number of targets yielded $10,554$ cases. For each case, we (both automatically and manually) coded the structure of noun phrases as described in (3).

(3)
- a. HASARTICLE: TRUE iff the given construction has an article (e.g., *a(n)*, *the*), demonstrative (e.g., *this, that*), or quantifier (e.g., *some, many*), FALSE otherwise (Abbreviated as HASART).
- b. HASPOSSESSIVE: TRUE iff the given construction has a possessive marker (e.g., *-'s, of, my*), FALSE otherwise (Abbreviated as HASPOSS).

We automatically classified all instances into four categories using the two criteria defined in Table 3. We coded these characteristics as independent to match the Löbner's separate classification. In coding the structural attributes of noun phrases, we also extracted the head noun of each noun phrase to lemmatise automatically.

Table 4: Initial results of analysis

NP type	Token Freq (%)		Type Freq (%)	
Cx1	844	(0.080)	485	(0.103)
Cx2	810	(0.077)	532	(0.113)
Cx3	3160	(0.299)	1367	(0.291)
Cx4	5746	(0.544)	2310	(0.492)

To conduct collostrustruction analysis, we used `Coll.analysis 4.0` (Gries, 2022a), functions to perform a family of collostructional analyses on R (R Core Team, 2022)[6]. To visualise the obtained results, we used `ggplot2` (Wickham and Grolemund, 2016).

4 Results & Discussion

This section reports the results of the collostruction analysis and their interpretations, respectively. Section 4.1 reports frequent lemmas of each construction, and Section 4.2 discusses a usage-based interpretation of our analysis.

4.1 Frequent lemmas of each construction

Observed token frequency and type frequency are summarised as Table 4. As confirmed, the values of tokens correlate with those of type frequency.

The overall distributions of $\Delta P_{cx \to w, w \to cx}$ are visualised as Figure 1. All lemmas are coloured by the construction types in which they occur. As can be confirmed in the figure, the lemmas of Cx4 are concentrated on the left part of the plot, meaning that most of the lemmas are not attracted by the NP construction. Figure 2 visualises the separated scatterplot excluding the cases with low frequency.

Figure 3 visualises the dispersions of construction frequency as boxplots. Each x-axis represents the logged frequency, and each y-axis represents the cumulative percentage. As can be confirmed on the plot, non-possessive constructions (i.e., Cx3, Cx4) have larger dispersions since they are more likely to be used.

In the following subsubsections (4.1.1–4.1.4), we report the results of collostruction analyses and their interpretations. As confirmed in Figure 1, some lemmas' $\Delta P_{w \to cx}$ are concentrated on 1.00 because of the low frequency. If a lemma occurs

[6] `Coll.analysis 4.0` is an updated script to perform collostrustruction analyses. Out of three collostrustruction analyses, we executed collexeme analysis to observe the attraction/repelling relations between head lemmas and instantiated noun phrases. The latest version of analysis was implemented in recent works (Gries, 2019, 2022c, 2023b).

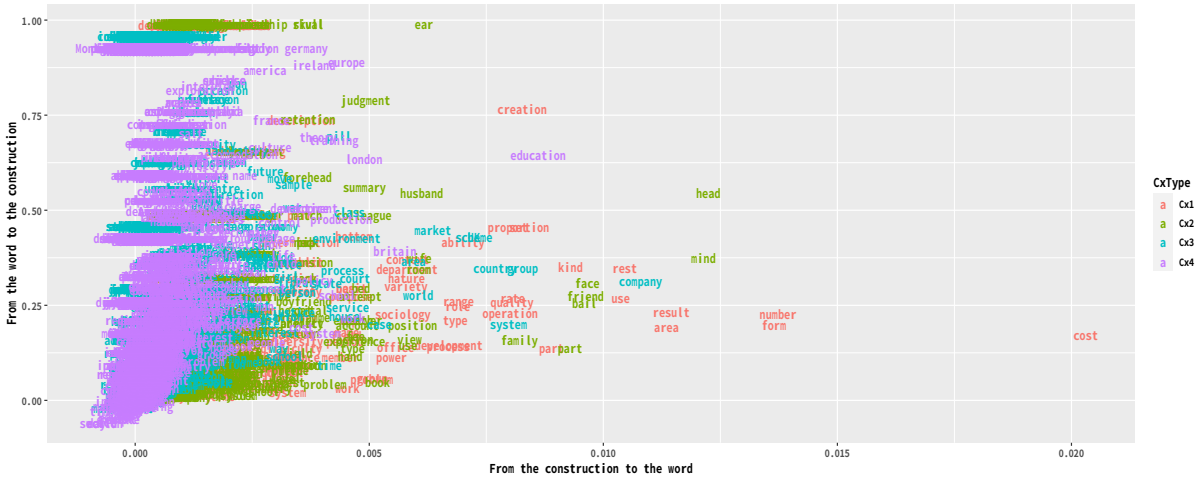

Figure 1: The scatterplot of all lemmas

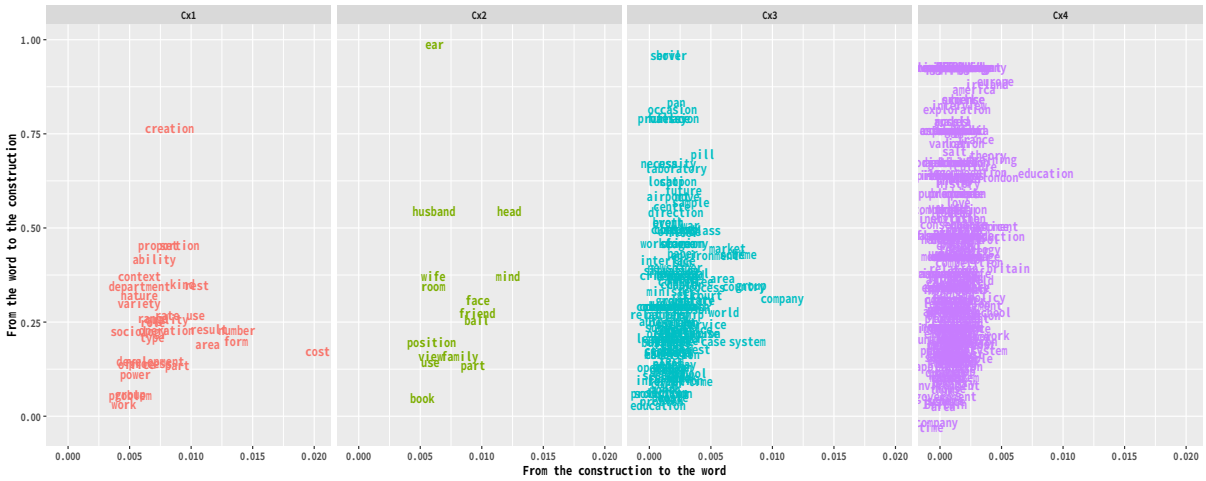

Figure 2: The scatterplot of each Cx type (Freq ≥ 5)

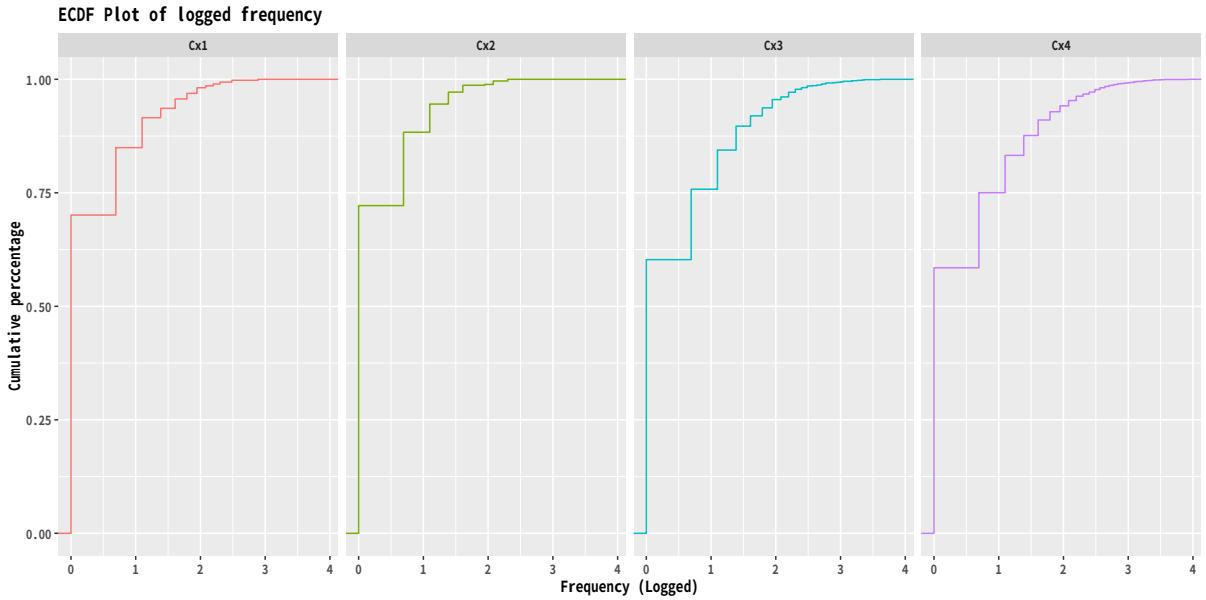

Figure 3: Dispersions of frequency in each construction

Table 5: Frequent lemmas of Cx1 (HASART, HASPOSS)

	$\Delta P_{cx \to w}$	$\Delta P_{w \to cx}$
1	cost	ambit
2	number	fraction
3	form	instinct
4	result	island
5	area	rear
6	rest	remains
7	use	reversal
8	kind	abandonment
9	part	advent
10	creation	amenity

Table 6: Frequent lemmas of Cx2 (HASPOSS)

	$\Delta P_{cx \to w}$	$\Delta P_{w \to cx}$
1.	head	ear
2.	mind	rival
3.	face	skull
4.	friend	adviser
5.	bail	championship
6.	part	humour
7.	family	ordeal
8.	ear	waist
9.	husband	accumulation
10.	wife	adherent

Table 7: Frequent lemmas of Cx3 (HASART)

	$\Delta P_{cx \to w}$	$\Delta P_{w \to cx}$
1	company	boil
2	group	server
3	system	attic
4	country	corridor
5	scheme	gallery
6	UK	kernel
7	market	module
8	world	peace
9	area	postcard
10	case	archive

only once, it is evaluated to attract the construction type in which it occurs. Since Glass's approach only captured the percentage of possessive construction that a given word can appear, her analysis could have evaluated these words as being highly relational. Figure 2 separates the distributions of $\Delta P_{w \to cx, cx \to w}$ by limiting the frequency to 5 or more to make the plot more visible.

Since the number of analysed lemmas is too large to interpret, we extracted the top ten words of each ΔP (i.e., the total of 20 lemmas) for each construction. As will be clear in the following discussions, the obtained results concord with previous studies.

4.1.1 Cx1

The summary of frequent lemmas of Cx1 is summarised as Table 5. Frequent lemmas include relational nouns (e.g., *cost, kind, part, fraction, rear. ...*) and eventive nouns (e.g., *creation, abandonment*).

Some lemmas (e.g., *number, kind, ...*) are more likely to be expressed as the part of multiword expressions (e.g., *"the {number, kind, ...} of ..."*). Lemmas that occur as heads of such multiword expressions (Sag et al., 2002) are often relational since their denotation must be determined in relation to the arguments they take (Grimshaw, 1990). As will be clear from the characteristics of Cx2, denotations of these terms must be determined in relation to other entities (e.g., *the cost of an increased mortgage*), unlike body part terms.

4.1.2 Cx2

The summary of frequent lemmas of Cx2 is summarised as Table 6. Most of the lemmas with high ΔP are relational (e.g., *friend, part, husband, ...*), and many of them are body-part terms (e.g., *head, face, ear, ...*).

The result is consistent with that of Glass (2022). Since Glass shows that frequencies of kinship and body-part terms rise as they licence the possessive constructions. However, unlike Cx1, these body part terms are more likely to collocate with pronoun possessives (e.g., *her head, his face*).

4.1.3 Cx3

The summary of frequent lemmas of Cx3 is summarised as Table 7. In contrast to Cx1 and Cx2, frequent lemmas are more likely to be SORTAL (e.g., *company, group, boil, server, ...*) in Löbner's sense.

As discussed, frequent lemmas of Cx3 are more likely to be SORTAL in contrast to Cx1 and Cx2. However, it does not mean that the identification of relational nouns is always successful since Table 7 includes relational nouns such as *attic* and *corridor*. Though denotations of these words must be determined in relation to the building that contains such parts, they are likely to be expressed without possessive phrases. Though the values of $\Delta P_{w \to cx}$ can be effected by the frequencies, this result suggests

Table 8: Frequent lemmas of Cx4 (Bare nouns)

	$\Delta P_{cx \to w}$	$\Delta P_{w \to cx}$
1	education	germany
2	britain	fiction
3	london	custody
4	work	monopoly
5	europe	warming
6	production	correspondent
7	school	agriculture
8	training	Lister
9	system	california
10	theory	japan

the effectiveness of collostructional analysis since it can distinguish the two directions of attractions.

4.1.4 Cx4

The summary of frequent lemmas of Cx3 is summarised as Table 8. Most of the frequent lemmas are proper nouns (e.g., *Britain, London, Germany, California*) and abstract nouns (e.g., *education, work, fiction, custody*). As already pointed out, the type frequency of Cx4 is the largest of all construction types since they subsume many proper nouns.

Frequent lemmas of Cx4 include both proper nouns and abstract nouns. The former class correspond to INDIVIDUAL, and the latter to SORTAL. Since we did not annotate if the given noun is a multiword expression (e.g., the number of students), or a proper noun (e.g., Ritsumeikan University, Kyoto University of Foreign Studies), some frequent lemmas could be a part of multiword proper nouns. Though this result can be easily predicted considering the grammatical constraints of English grammar, it can be said that Löbner's characterisation of this construction should be revised to identify the uniqueness of a nominal expression. It could be effective to include features used for mass/count distinctions (Jackendoff, 1992).

4.2 Typical instances of noun phrases

Based on the values of $\Delta P_{\{cx \to w, w \to xc\}}$, our collostructional analysis allows us to identify the typical instances of each NP construction. We identified typical instances by combining the values of ΔP, and argue that our approach can smooth out the skewed frequency of constructions.

In a usage-based model of language (Langacker, 1999), language knowledge is seen as a vast inventory of actual uses (Taylor, 2012). The usage-based conception of language aligns well with assumptions of construction grammars (Hoffmann, 2022, 266–269). By taking frequency into account, analysts can consider the typicality of the construction under analysis.

We identified the prototypical instances of NP constructions by comparing every value of $\Delta P_{\{cx \to w, w \to cx\}}$ with the respective medians. For instance, the median of $\Delta P_{cx \to w}$ in Cx1 is 0.001. We compared if every value of $\Delta P_{cx \to w}$ is larger than 0.001. By repeating the process to each direction of association and each construction type, we identified the typical instances of each construction as a form of conjunction (i.e., The given value of $\Delta P_{cx \to w}$ is larger than the median of $\Delta P_{cx \to w}$ AND the given value of $\Delta P_{w \to cx}$ is larger than the median of $\Delta P_{w \to cx}$).

As a result, we obtained the typical nouns that occur in each construction type, which is summarised as Table 9. As discussed in Section 4.1, a value of $\Delta P_{w \to cx}$ can easily be effected by the low-frequent items. However, since we considered both directions, the effects of low-frequency items are smoothed out. Nouns in Table 9 roughly correspond to typical instances of each semantic class of Löbner (2011).

5 Conclusion

This paper sketched a collostructional approach to the uniqueness and relationality of nouns. By considering two kinds of attractions (i.e., attraction from the word to the construction and the construction to the word), we succeeded in describing the complex interactions of concept types and collocating constructions, which was consistent with previous studies.

Limitations

Type-shifting rules and construction grammar

Löbnerian semantic shift rules propose that each noun of a given semantic type can be transformed into any other type (Löbner, 2011, 313). While this approach may seem elegant from an algorithmic perspective, it is unclear whether it accurately represents our linguistic knowledge. Our collostructional analysis allows establishing the prototype condition for nominal constructions (e.g., [$_{\text{NP}}$ N_1's N_2]), which aligns with the principles of the usage-based model in cognitive linguistics (Langacker, 1988, 1999).

Table 9: Typical nouns of each construction (Sorted by raw frequency)

	Cx1	Cx2	Cx3	Cx4
1	number	head	group	germany
2	form	mind	country	fiction
3	result	face	scheme	custody
4	area	friend	UK	monopoly
5	rest	bail	market	warming
6	use	ear	area	correspondent
7	kind	husband	class	agriculture
8	creation	wife	environment	Lister
9	proportion	room	process	california
10	set	judgment	pill	japan

Interpreting strength of determining the relationality and uniqueness

Though employing collostructional analysis led to finer-grained descriptions of the uniqueness and relationality of given nouns, the direction of attractions must be considered more carefully. Since values of ΔP are easily effected by the frequency of the key expression, analysts must devise a way to evaluate the collostructional strength to determine the strength of relationality and uniqueness. However, we could not implement such measures.

Correspondence between currently available semantic classes of nouns

Glass discusses the correspondence between the conventional semantic class (e.g., HUMAN, OCCUPATION) and the frequency of possessive constructions. Since this paper aimed to refine and extend the computational aspects of Glass (2022), we did not explore the semantic correspondence. Additional annotations must be carried out to accurately compare our results with Glass (2022), and its strategy must be assessed carefully.

Ethics Statement

This article does not contain any studies with human or animal subjects performed by any of the authors.

Data availability

The annotated data, the result of collostructional analysis, and the R codes used for the analysis are available on `https://osf.io/3knzy/?view_only=a2c0620d2ee241f08ae69b1cb153e733`.

Acknowledgements

We wish to thank the two anonymous reviewers for their helpful comments. All remaining errors are ours.

References

Jennifer A. Asmuth and Dedre Gentner. 2005. Context sensitivity of relational nouns. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, pages 163–168.

Chris Barker. 2011. Possessives and relational nouns. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics: An International Handbook of Natural Language Meaning*, pages 1109–1130. Mouton de Gruyter, Berlin.

Alastair Butler. 2022. *The Treebank Semantics Parsed Corpus (TSPC)*. Hirosaki University, Aomori, Japan. (accessed 9 June 2023).

Jos de Bruin and Remko Scha. 1988. The interpretation of relational nouns. In *26th Annual Meeting of the Association for Computational Linguistics*, pages 25–32.

Lelia Glass. 2022. Quantifying relational nouns in corpora. *English Language & Linguistics*, 26(4):833–859.

Stefan Th. Gries. 2019. 15 years of collostructions: Some long overdue additions/corrections (to/of actually all sorts of corpus-linguistics measures). *International Journal of Corpus Linguistics*, 24(3):385–412.

Stefan Th. Gries. 2022a. *Coll.analysis 4.0. A script for R to compute perform collostructional analyses*.

Stefan Th. Gries. 2022b. Toward more careful corpus statistics: uncertainty estimates for frequencies, dispersions, association measures, and more. *Research Methods in Applied Linguistics*, 1(1):100002.

Stefan Th. Gries. 2022c. What do (some of) our association measures measure (most)? association? *Journal of Second Language Studies*, 5(1):1–33.

Stefan Th. Gries. 2023a. Overhauling collostructional analysis: Towards more descriptive simplicity and more explanatory adequacy. *Cognitive Semantics*, 9(3):351–386.

Stefan Th. Gries. 2023b. Quantitative corpus methods of cognitive semantics/linguistics. In Thomas Li, editor, *Handbook of Cognitive Semantics*, pages 328–350. Brill, Leiden.

Stefan Th. Gries and Anatol Stefanowitsch. 2004a. Covarying collexemes in the into-causative. In Michel Achard and Suzanne Kemmer, editors, *Language, Culture, and Mind*, pages 225–236. CSLI, Stanford.

Stefan Th. Gries and Anatol Stefanowitsch. 2004b. Extending collostructional analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics*, 9(1):97–129.

Jane Grimshaw. 1990. *Argument Structure.* MIT Press, Cambridge, Mass.

Martin Hipert. 2014. Collostructional analysis: Measuring associations between constructions and lexical elements. In Dylan Glynn and Justyna A. Robinson, editors, *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, pages 391–404. John Benjamins, Amsterdam & Philadelphia.

Thomas Hoffmann. 2022. *Construction Grammar*. Cambridge University Press, Cambridge.

Thomas Hoffmann and Graeme Trousdale, editors. 2013. *The Oxford Handbook of Construction Grammar*. Oxford University Press, Oxford.

Ray Jackendoff. 1992. Parts and boundaries. In Beth Levin and Steven Pinker, editors, *Lexical & Conceptual Semantics*, pages 9–54. Blackwell, Oxford.

Ronald W. Langacker. 1988. A usage-based model. In Brygida Rudzka-Ostyn, editor, *Topics in Cognitive Linguistics*, pages 127–161. John Benjamins, Amsterdam & Philadelphia.

Ronald W. Langacker. 1999. A dynamic usage-based model. In Michael Barlow and Suzanne Kemmer, editors, *Usage Based Models of Language*, pages 1–63. CSLI Publications, Stanford.

Sebastian Löbner. 2011. Concept types and determination. *Journal of Semantics*, 28(3):279–333.

Malvina Nissim. 2004. Lexical information and choice of determiners. In Ji-yung Kim, Yury A. Lander, and Barbara H. Partee, editors, *Possessives and Beyond: Semantics and Syntax*, pages 133–52. Brill, Leiden.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Computational Linguistics and Intelligent Text Processing*, pages 1–15.

Anatol Stefanowitsch and Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics*, 8(2):209–243.

Anatol Stefanowitsch and Stefan Th. Gries. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory*, 1(1):1–43.

John R. Taylor. 2012. *The Mental Corpus: How Language is Represented in the Mind*. Oxford University Press, Oxford.

Hadley Wickham and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Inc., California.