Constraints on Linking Element Choice in German Nominal Compounding: A Large-Scale Corpus Study

Maksim Shmalts
Department of Linguistics
University of Tübingen
maksim.shmalts@uni-tuebingen.de

Abstract

The N+N compound class is the largest and the most productive class of compounds in German. A significant number of N+N compounds insert a so-called linking element from a large inventory. The linker choice is notoriously irregular; instead of rules, it is governed by a set of constraints that can only limit this choice based on morphological, phonological, sometimes semantic and lexical properties of the first constituent. While constraints on linking element choice in German nominal compounding are extensively researched and well-documented, no large-scale corpus study has ever been reported on the subject of their empirical application. The present work aims at filling in this gap by conducting an extensive corpus study on potential and actual applicability of these constraints. The study summarizes 64 constraints collected from the relevant literature and obtains applicability statistics for 39 of them over 280k+ German N+N compounds. The study both confirms most of the evidence from previous literature and suggests novel evidence on German nominal compounding. It additionally highlights the importance of structured linguistic data for large-scale empirical studies.

Keywords: compounding, linguistic constraints, corpus study, German morphology

1. Introduction

Compounding is a process of word formation broadly defined as concatenation of two constituents with or without adding a linking element (also called linker) in between: Buch '*book*' + Deckel '*cover*' → Buchødeckel '*book cover*' but Buch '*book*' + Regal '*shelf*' → B**ü**ch**er**regal '*bookshelf*'. The first constituent becomes the modifier, and the second the head of the resulting composite. In German, compounding is highly productive: thus, Baroni et al. (2002) report that in the APA newswire corpus (28M+ tokens), 47% of word types are compounds. Biconstituent nominal (N+N) compounds in German form the largest class of compounds (Ortner and Müller-Bollhagen, 1991, 53; Scherer, 2012; Schlücker, 2023). A sizable number of such compounds inserts an explicit linking element between the constituents. While German features "an unusually high number of different linking elements" (Schäfer and Pankratz, 2018), only one of them is grammatically correct for a given compound[1]. What makes the phenomenon notoriously complicated is the fact that it is often arbitrary which exact linker from the extensive inventory is the correct one. This choice is non-deterministic and cannot be predicted confidently from any properties of the constituents alone. On the contrary, it is governed by a large set of (irregular) *constraints* that can only limit this choice based on morphological, phonological, sometimes semantic and even lexical properties of the first constituent (Eisenberg, 2013, 227; Kürschner; Libben et al., 2009; Schäfer and Pankratz, 2018); the second constituent is argued to have no effect on linker choice[2] (Kürschner; Ortner and Müller-Bollhagen, 1991, 56).

Constraints on nominal compounding[3] in German have been extensively researched and well-documented in the linguistics literature over the last decades. Their irregularity and diversity is well-known and generally agreed upon. And yet, despite the high productivity of nominal compounding in German and the high complexity of the corresponding constraint space, no *large-scale* corpus study has ever been reported on the subject of *empirical* constraint application.

Our study serves as an initial attempt to investigate the empirical application of nominal compounding constraints in German quantitatively. More specifically, we focus on their *coverage* and *regularity*. By the coverage of a constraint, we mean the share of the units that satisfy the linguistic conditions for this constraint to be *potentially applicable*. In our terminology, such units are also referred to as units *covered* by this constraint. The regularity of a constraint is then the share of the units covered by this constraint that actually conform to it. In line with the introduced terminology, we call a constraint *actually applicable* to the units that conform to it, or simply say it *applies* to them. For instance, a dummy constraint (0) formulated

[1] Except for a small number of so-called doubtful cases when two variants coexist in language, see Nübling and Szczepaniak (2013).

[2] See exceptions in section 2.2.

[3] To clarify: by 'nominal compounding constraints', we mean constraints on choice of the linking element in N+N compounds.

Proceedings of the Workshop on Structured Linguistic Data and Evaluation, pages 14–38

as "First constituents that are constituted by nouns of the A class tend to attach -*x*-." will be potentially applicable to dummy compounds $A\mathbf{x}B$ and $A\mathbf{y}C$, but not to $C\mathbf{x}A$, and will be actually applicable only to $A\mathbf{x}B$. The formal definition of coverage and regularity follows in section 4.2.

We compute the coverage and the regularity of the constraints programmatically over a modification of a compound dataset originating from (Schäfer and Pankratz, 2018). Access to *strictly structured* linguistic data becomes a prerequisite for automated computation: only well-defined structures over the dataset can enable a uniform and reliable programmatic access to the target linguistic properties as defined by constraint applicability conditions. CELEX2 (Baayen et al., 1995) becomes the natural choice for us: this lexical database stores various linguistic information about its entries in a table-like format perfectly suitable for automated processing.

The present study makes two contributions to the developments in German nominal compounding studies: 1) Constraints on linker choice in German N+N compounds are collected from a number of relevant scientific sources, summarized and classified. 2) A corpus study is conducted on the matter of empirical application of the collected constraints[4]. The combined effect of these two contributions results in a systematic constraint space. Each constraint in the constraint space comes with the information about the type of linguistic information it employs (morphological, phonological, etc.), and, whenever possible, the measures of its coverage and regularity over 280k+ N+N compounds. This development is relevant both to descriptive studies in theoretical linguistics on German morphology and as a knowledge base/graph for novel compound generation with LLMs. The latter may be useful for evaluating linguistic competencies and reasoning abilities of LLMs and may increase the quality of generated compound datasets.

The paper is structured as follows. Section 2 summarizes principles of German nominal compounding. Section 3 gives an overview of previous work concerning German nominal compounding (in particular, compounding constraints). Section 4 introduces the methodology and the course of our empirical study. Finally, the results of the study are discussed and summarized in sections 5 and 6.

2. On Nominal Compounding in German

The N+N compound class is the largest and the most productive class of compounds in German. We consider solely N+N compounds because compounding in German conforms to binary recursion (Eisenberg, 2013, 219; Schäfer, 2018, 226-227), and so any compound can be decomposed into a chain of binary structures. While German (and Germanic in general) has a few constraints that specifically target compounds with three and more constituents[5] (see Krott et al., 2004; Kürschner; Wegener, 2005), we exclude them from the present study due to their small number. Furthermore, N+N compounds is the class where the explicit linking elements occur most often (Schlücker, 2023).

2.1. Linking Elements

The inventory of linking elements in German includes zero linker -*ø*-, the following explicit linkers: -*s*-, -*es*-[6], -*(e)n*-, -*e*-, -*"e*-[7], -*(")er*-[8], -*"*-, -*(e)ns*-, and a number of idiosyncratic linkers for loanwords. The latter — as well as linkers causing deletion of final stem segment(s) — are disregarded in the current study due to their insignificant frequency.

The correct linker is often not inferrable with certainty from any properties of the first constituent. The arbitrariness of choice may be so high that some nouns can attach different linkers in different compounds, e.g., Land '*state*' + Kreis 'district' / Regierung 'government' / Spiel 'match' → Land*ø*kreis '*regional district*' / Land**es**regierung '*regional government*' / L**ä**nd**er**spiel '*international match*'. Augst (1975, 134) finds 9.3% of entries have records with two different linkers and almost 1% of entries with three or four different linkers in a corpus of roughly 4k compounds. Rather than predict the correct linking element for a given noun, one can only limit the group of acceptable linkers, guided by a set of (irregular) constraints.

2.2. Constraints on Linker Choice

Constraints on nominal compounding in German are defined almost exclusively over linguistic information of the first constituent. Ortner and Müller-Bollhagen (1991, 56) document the single case

[4] Excluding semantic and lexical constraints, see more in section 4.4.

[5] Thanks to one of the anonymous reviewers for bringing that to our attention.

[6] -*s*- and -*es*- are not notated as -*(e)s*- as they are not allomorphic (-*es*- is isolated, see Eisenberg, 2013, 229; Kürschner, 2010), in contrast to -*(e)n*- and -*(e)ns*-.

[7] *"* signals that the linker triggers umlauting of the stem vowel of the first constituent.

[8] To the best of our knowledge, the -(")er- linker triggers umlauting necessarily whenever the stem vowel may undergo it. We therefore do not distinguish between -er- and -"er- (as opposed to -e- vs. -"e-) as cases like Kind**er**jacke '*children's jacket*' may be accounted for by assuming that -(")er- tries to apply umlauting but the target vowel cannot not be umlauted. We hold the same assumption against the -(")er plural marker.

when the properties of the second constituent contribute equally to linker choice, see constraint (39) in the list of example constraints below. Apart from that, Eisenberg (2013, 227) suggests that sometimes, the semantic relation between the two constituents may play a definitive role in linker choice. This claim is supported by evidence provided by Koliopoulou (2014), also Neef and Borgwaldt (2012), who independently formulate constraint (36).

The rest of the constraints refer solely to the properties of the first constituents and are based on various types of linguistic information. The literature documents morphological, derivational[9], phonological, semantic, lexical, and mixed constraints. Below is an example list of constraints collected from Eisenberg (2013, 227-230) and Ortner and Müller-Bollhagen (1991, 56-57, 83, 89-94, 109).

(7) [morphological] First constituents that are constituted by nouns that build the plural form with *-(")er* mostly attach *-ø-* or *-(")er-*: Buch '*book*' + Deckel '*cover*' → Buch**ø**deckel '*book cover*' but Kind '*child*' + Alter '*age*' → Kind**es**alter '*childhood*'

(22) [derivational] First constituents that are constituted by deadjectival feminine nouns with a schwa suffix almost always attach -(e)n- or a zero linker; each of the two linkers is preferred in about the same number of cases: Hitze '*heat*' + Welle '*wave*' → Hitze**ø**welle '*heatwave*' but Liebe '*love*' + Brief '*letter*' → Liebe**s**brief '*love letter*'

(36) [semantic] Copulative compounds insert *-ø-* regularly (by copulative compounds are understood compounds in which the two constituents do not exhibit a clear modifier-head relation): Dichter '*poet*' + Komponist '*composer*' → Dichter**ø**komponist '*poet-composer*'

(39) [semantic, lexical] Compounds whose first constituent designates a person and whose second constituent is one of 'Mann', 'Frau', 'Leute', 'Tochter', 'Gattin', or 'Witwe' tend to insert *-s-*: Reiter '*rider*' + Mann '*man*' → Reiter**ø**mann '*horseman*'

(46) [morphological, phonological] Almost all feminine nouns that constitute first constituents that attach -s- are polysyllabic: Arbeit '*work*' + Tag '*day*' → Arbeit**s**tag '*working day*'

As will be demonstrated in sections 4.1 and 5, nominal compounding constraints in German are highly diverse in terms of their regularity and coverage, as well as regarding the type of linguistic properties they target.

[9] To the best of our knowledge, this term is not well-established. By derivational constraints we mean those that target classes of nouns with specific derivational structure, e.g., nouns with specific affixes, deverbal nouns, etc.

3. Previous Research

Linking elements in (N+N) compounds have been a subject of interest in German morphology for decades. Numerous works address different aspects of their development, functionality, and linguistic status.

Nübling and Szczepaniak (2013), Wegener (2003), Wegener (2005) investigate the origin and the diachronic development of the linking elements. They identify two sources of German linking elements: Germanic primary suffixes and later genitive markers. By addressing the process of their lexicalization and reanalysis, the authors explain the restrictions on linker choice in contemporary German from a historical perspective. In particular, they suggest that plural markers *-(e)n*, also *-e*, *-"e*, *-(")er* originate from the primary suffixes, and the homonymous linking elements evolve either from these plural markers (Nübling and Szczepaniak, 2013), or in parallel to them (Wegener, 2003). This way, these works motivate the fact that these linkers appear exclusively with nouns of the corresponding plural[10].

A larger share of the literature body reports contemporary limitations on linker choice without focus on their origin. For example, Eisenberg (2013, chapter 6.2), Ortner and Müller-Bollhagen (1991, chapter A.5), Schäfer (2018, chapter 8.1) expound general principles of German (nominal) compounding, each covering acceptability conditions of most of the investigated linkers. Numerous works are dedicated specifically to the *-s-* linker since it is the only productive linker not bound to any inflectional class: Libben et al. (2009), Kopf (2017), Kürschner, Nübling and Szczepaniak (2008).

Finally, a notable branch of research is oriented towards investigating various analogical effects associated with compounding in German. Thus, Krott et al. (2007) find that choice between *-s-*, *-(e)n-*, and *-ø-* is biased towards the most frequent linker occurring with the given first constituent in other compounds, with a statistically significant effect. With different experimentation course, Neef and Borgwaldt (2012) arrive at similar results and confirm the left constituent bias. They also suggest that building novel compounds is subject to inter-speaker variability to a certain degree. On another note, Schäfer and Pankratz (2018) provide evidence that linkers identical to the plural markers may serve a clue to a plural or collective meaning of the first constituent in the compound.

[10] For -(e)n- and -e-, Nübling and Szczepaniak (2013) also identify edge cases not related to the plural suffixes.

All these findings are of undeniable value and importance. Combined, they provide an exhaustive yet granular overview of the constraints on linker choice in German. However, the absolute majority of these works have a significant fallback: while noting the infamous irregularity of these constraints, they fail to deliver its empirical measures. This results in quite vague constraint definitions that sometimes hinder building an adequate idea of a certain constraint. A few examples follow[11]: "If the stem [of a feminine noun] is monosyllabic, the *-en-* linker is *sometimes* inserted, *sometimes* not." (Eisenberg, 2013, 229); "There is an *observable tendency* to insert *-e-* after nouns designating animals [...] with many counterexamples." (Nübling and Szczepaniak, 2008, 11); "Strong and mixed masculine and neuter nouns *often but by no means always* insert *-s-* or *-es-*[12]" (Schäfer, 2018, 229).

Clearing up such vague definitions would only be possible by the means of a dedicated corpus study. However, only targeted corpus studies covering single constraints have been made so far. Kopf (2017), Kürschner, and Nübling and Szczepaniak (2008) aim at investigating the increased probability of *-s-* after prefixed deverbal nouns. They confirm that prefixed deverbals attach *-s-* in a majority of cases: 67.5%, 76.5-78%, and 85% (with unstressed prefix) / 36% (with stressed prefix), respectively. Going further, Nübling and Szczepaniak (2008) and Wegener (2005) address the relation between the probability of *-s-* and the sonority of the last segment of the noun. They find that *-s-* occurs after plosives in 15-20% and after sonorants in 1.8-4.7% of cases, and it never occurs after a full vowel.

To the best of our knowledge, no attempts have been reported yet of extensive corpus studies that would cover the entire space of nominal compounding constraint or at least a large part of it. We therefore aim at filling in this gap by conducting a novel large-scale corpus study on the application of nominal compounding constraints in German.

4. Empirical Study

The present study was conducted in four steps. We briefly summarize them here, and introduce them in more detail in the corresponding subsections 4.1-4.4 below.

Step one fixes the definitions and linguistic status of constraints from the relevant literature.

Step two ensures reproducibility of the study. It establishes formalizations concerning potential/actual applicability of the collected constraints and calculation of their coverage and regularity.

The third step prepares the data for the corpus experiments. It obtains a massive collection of N+N compounds taking a compound dataset DeCOW16AX-comps originating from Schäfer and Pankratz (2018) as a baseline. It then derives a custom noun database from CELEX2 (Baayen et al., 1995) — a large lexical database for English, German, and Dutch that stores various linguistic properties of its entries (lemmata and wordforms). Additionally, it retrieves word counts of the entries of the two obtained modifications from DeReKo (IDS, 2026) — the largest collection of text corpora in contemporary German.

Step four conducts the corpus study itself. It estimates the empirical applicability of the collected constraints over our modification of DeCOW16AX-comps following the procedural formalization from the second step.

The developments of the empirical study — including the collected constraint definitions and supporting information, the compound dataset, and the relevant codebase — are published under `https://github.com/maxschmaltz/DieKonstraints/tree/comp_constr_corp_study_mar26`. Please refer to the repository's README to navigate the project files.

4.1. Step One: Constraint Acquisition

To obtain linker choice constraints reported in the linguistics literature so far, we studied a number of scientific works dedicated to German (nominal) compounding. We aimed at keeping the collection of scientific sources diverse in terms of their format, purpose, object of research, and date of publication. In case of discrepancies, we did not take sides with one or another analyses and kept both sources, assuring a wide spectrum of linguistic perspectives. The resulting body of literature spans over 30+ years and includes chapters of fundamental books on German morphology in general and on (nominal) compounding in particular, survey-like papers reporting generalizations of German compounding and linker choice, as well as more targeted papers addressing specific aspects of compounding in German, such as their functionality and grammatical status, restrictions on their insertion, etc. To adopt a broader perspective, we also involved works dedicated to the origin of linking elements, analogical effects in compounding, and comparative studies on linking elements in Germanic. The resulting set of works includes 18 sources and is provided in Appendix A.

From these sources, we extracted claims that described restrictions on, or patterns of, choice of the linking elements in N+N compounds. We find no critical contradictions across the works. The

[11] Whenever necessary, translated by the author of the present study. In italics, we highlight the vague spots.

[12] The original work reads *-(e)s-* but as mentioned before, we distinguish between these two linkers.

only disagreements concern regularities or linguistic nature of a few constraints. These are of no significance to us, since we obtain our own statistics of constraint application. Otherwise, there is an ongoing debate on morphological (Kürschner) vs. phonological (Nübling and Szczepaniak, 2008) conditioning of *-s-*. We simply formulate both options as two separate constraints for further verification.

In total, we summarized 64 constraints. Apart from the default constraint striving to assign a zero linking element to any given compound, two main categories of constraints could be established. The first category is prevalent with 41 constraints. It limits acceptable linkers for a given noun class ('*A* can attach either *x* or *y*'), as exemplified by constraints (7), (22), (36), (39) above; we therefore call it *P2L* (properties to linkers). The second category (22 constraints) goes the opposite way: it puts restrictions on the noun classes that can attach a given linker ('*x* can only be attached by *A* or *B*'), as exemplified by constraint (46); correspondingly, we name it *L2P*. Table 2 exhibits constraint types by level(s) of accessed linguistic information[13]. Both categories actively employ all types of linguistic information, involving prevalently morphological but to a noticeable extent derivational, phonological, and semantic properties. That observation is in line with the fact that linker choice is highly variable and is affected by a large variety of factors.

Fifty-nine out of 64 constraints are defined over the properties of the first constituent. Note that the 22 L2P constraints which all belong to that group additionally require insertion of a certain linker to be potentially applicable. For instance, constraint (46) covers not any feminine first constituent but only those that attach *-s-*. From the remaining five constraints, one depends on the properties of both constituents, three on the semantic relation between them, and the last one is the default constraint that is potentially applicable to any given constituent combination. This finding confirms the general agreement that the linker choice depends almost exclusively on the properties of the first constituent. In the following, the component(s) of a compound accessed by a constraint will be called the target *item* of this constraint in the compound. For most constraints, the first constituent of a compound or the combination of the first constituent and the linker is the target item. For the five remaining constraints mentioned above, the combination of both constituents is the target item.

The full list of constraint definitions, properties, and supporting information (P2L/L2P type, employed linguistic information, target item, examples, and counterexamples) is available in Appendix E.

Type	P2L	L2P
morphological	6	6
+ derivational	1	1
+ phonological	1	0
+ phonological	2	1
+ semantic	3	1
+ lexical	0	1
+ semantic	8	1
+ lexical	0	1
derivational	9	2
+ semantic	2	0
phonological	5	4
semantic	2	0
+ lexical	1	0
lexical	1	4

Table 1: Types of nominal compounding constraints by employed linguistic information. Mixed types are marked by indentation. Mid-grey marks types of zero, and light-grey of up to three constraints.

4.2. Step Two: Procedural & Mathematical Formalizations

Every constraint definition consists of two components. The first component declares the linguistic properties of the constraint's target item that determine potential applicability of this constraint to an arbitrary compound formed from it. For most constraints, the first component is effectively the description of the desired properties of the first constituent. The second component specifies a set of acceptable linkers that are expected to appear in a compound built with this item if the constraint applies. For instance, constraint (7) (target item: first constituent) may be represented[14] as `[N1-pl = (")er]` → `[-(")er-, -Ø-]`. Constraint (39) (target item: the combination of both constituents) can similarly be formulated as `[N1 is pers][N2 ∈ {Mann, Frau, ...}]` → `[-Ø-]`. Any constraint can then be abstracted as `[cond 1]...[cond` n`]` → `[link 1, ..., link` m`]`. Then, the formal tests for potential and actual applicability are trivial. For an arbitrary constraint (X) and an arbitrary N+N compound C formed from item I, it holds: 1) (X) is potentially applicable to C if for $i = (1, ..., n)$, **all** `[cond` i`]` defined over I are true; 2) (X) applies to C if (X) is potentially applicable to C and

[13] Derivational constraints defined over morphologically restricted classes are not additionally marked as morphological if they target nouns with suffixes that determine morphological properties of the derived noun (e.g., suffix *-heit* always produces weak feminines).

[14] The following abbreviations are adopted for these examples: `N1` and `N2` stand for the first and the second constituent, respectively, `pl` stands for 'plural marker', `pers` — for 'person designation', `cond` — for 'condition', `link` — for 'linker'.

if for $j = (1, ..., m)$, C inserts **any** of `[link j]`[15]. Please note that since all `[cond i]` are defined over the properties of I, potential applicability of (X) to C and to further compounds built from I will be identical. This is not the case with actual applicability which is determined by the linker in a specific compound. For the L2P constraints, the two constraint definition components and, correspondingly, the tests are swapped.

When calculating the statistics of the empirical constraint application, we will distinguish between *type coverage and regularity* and *item coverage and regularity*[16]. This distinction results from calculating the application statistics over unique compounds in the dataset or over unique items (mostly first constituents) that form these compounds, respectively. A characteristic example may be helpful to motivate this distinction. As discovered after conducting the corpus study, constraint (22) (target item: first constituent) is potentially applicable to 2798 compounds that are built from 32 unique first constituents. All 414 of 2798 compounds that violate (22) are built with one single first constituent Liebe '*love*' (Liebe**s**brief '*love letter*', Liebe**s**erklärung '*declaration of love*', etc.). That means that while $\sim 97\%$ unique items (31 out of 32 covered first constituents) always build compounds that conform to (22), they only account for $\sim 85\%$ of unique compounds. In other words, (22) is (almost) regular from the item perspective and is only a tendency from the compound perspective. As will be elaborated in section 5, there are further constraints that exhibit a similar gap. We therefore report two variants of coverage and regularity: the *type* variant and the *item* variant[17].

Formally, *type coverage* of a constraint (X) amounts to the ratio between the number of unique compounds to which (X) is potentially applicable as defined in the corresponding formal test, and the total number of unique compounds in the dataset. Since potential applicability of (X) to compounds built with the same item is identical, *item coverage* of (X) is defined simply. It equals the ratio between the number of unique items that build the compounds covered by X, and the total number of unique items found in the dataset.

Type regularity of (X) amounts to the ratio between the number of unique compounds to which (X) applies as defined by the respective formal test, and the number of unique compounds covered by (X). As opposed to potential applicability, (X) may (and often does) apply only selectively to different compounds built with the same item. To address this peculiarity, item regularity examines groups of compounds built with every unique item separately. Within each group, *group regularity* is calculated as the ratio between the number of unique compounds in the group that conform to (X) and the total number of unique compounds in the group. *Item regularity* of (X) is then the average over all group ratios. Item regularity δ of (X) thus conceptually means the following: on average, for any arbitrary item that satisfies linguistic conditions put by (X), $\delta \times 100\%$ of all compounds formed from this item conform to (X), regardless of the productivity of the item. Consequently, this method eliminates potential item productivity bias against type regularity exemplified by the first constituent Liebe with respect to constraint (22).

To ultimately clarify the provided mathematical formalizations, we demonstrate the calculation of the applicability statistics for constraint (22) in Appendix B. In the following, the item variant of coverage and regularity will be considered principal and will be reported first, followed by the type variant. This is motivated by the fact that, as demonstrated above, the item variant is not affected by the item productivity bias and is therefore more balanced and illustrative. Finally, we adopt the term of *combined coverage and regularity* for the contexts in which we simultaneously compare both types of the measures to some absolute value. Combined coverage/regularity $> \gamma$ of (X) means that both item and type coverage/regularity of (X) exceed γ. Correspondingly, combined coverage/regularity $< \gamma$ points that both types of coverage/regularity lie below γ.

4.3. Step Three: Data Acquisition

As pointed out in section 4, we employed three data sources for various purposes. DeCOW16AX-comps[18] was used to derive a massive N+N collection that served the corpus over which we assessed the empirical applicability of the constraints. CELEX2[19] was accessed during the experimental

[15] Note that if a constraint implicitly licenses further linkers, these are not accepted in our procedure, as these are exactly the irregular deviations that we want to record. For example, even though constraint (39) only *tends* to insert *-s-* — and so allows other linkers —, we only consider that it applies if attached is the *-s-* linker.

[16] Please note that this distinction is not the same as the customary distinction between type and token.

[17] In a dedicated branch of the study, we additionally calculated token coverage and regularity of the constraints that based on word counts of compounds. However, the values of the type and the token variants diverged insignificantly, and so the token variant was excluded from the present study.

[18] DeCOW16AX-comps is licensed under CC BY-NC 4.0 and are freely available for research purposes.

[19] CELEX2 is distributed under a custom license authorizing unrestricted usage for research purposes, see https://catalog.ldc.upenn.edu/docs/LDC96L14/celex.readme.html#Copyright. The data was obtained in compliance with this license.

phase to test for various linguistic information of compound constituents. Finally, DeReKo [20] supplied the word counts for the entries of our subsets of the first two. The preprocessing procedure of DeCOW16AX-comps depended on that of CELEX2, so we start the overview over data acquisition with the latter.

CELEX2 contains ∼51k German lemmata. For each lemma, it provides a rich set of linguistic information. Relevant for the present study are the records about POS, gender, inflectional paradigm, morphemic structure (both base word and affixes), phonetic transcription, and syllabification scheme of the lemmata. The high quality of annotations in CELEX2 minimized the number of preprocessing steps undertaken. Apart from a few minor preprocessing steps (mostly concerning data formatting), we simply extracted a subset of one-stem nouns that met two requirements: 1) they had the full set of the relevant records enlisted above; 2) their word count in DeReKo reached the minimum threshold of 50 occurrences. For word counts, we were querying a collection of DeReKo subcorpora compiled under the `copr-d` (Korpora aus Deutschland '*corpora from Germany*') namespace of the KorAP interface of DeReKo with a total of 15B+ tokens. This, together with the minimal creation year set to 1970, allowed us to avoid overt regionalisms or anachronisms. The frequencies were obtained via the KorAP OpenAPI endpoint (`https://korap.ids-mannheim.de/api/v1.0/openapi/`), the endpoint was accessed in early February 2026. The resulting sample of CELEX2 (henceforth also called CELEX-nouns) contains ∼12k nouns.

DeCOW16AX-comps supplied us with ∼22.3M SMOR (Schmid et al., 2004) analyses of German compounds with nominal heads. We took this collection as a starting point and in a few steps, dropped all records unsuitable for our study. The resulting modification that we called GeCoDB_v06 contains ∼280k N+N compounds featuring all variety of the linkers considered in the present study. Each record is a triplet identifying the two constituents and the linker of the compound. We only kept compounds where both constituents are present in CELEX-nouns — that ensured that we had access to the full set of required linguistic information for all of our compounds. To further ensure a high quality of the sample, we set a minimum word count threshold of 20 in DeReKo[21], and for its modifier to form at least ten compounds in GeCoDB_v06 with different heads.

[20] Most of the DeReKo subcorpora are licensed under a QAO-NC license. DeReKo may be used freely for scientific purposes.

[21] Word counts were collected under the same procedure that was utilized for CELEX2.

4.4. Step Four: Corpus Experiment

To be able to estimate constraint applicability over the GeCoDB_v06 corpus, we translated the procedural formalizations of the tests for potential and actual applicability from section 4.2 into Python code. The checker for potential applicability of a certain constraint accepts a compound and accesses the CELEX-nouns database to test for every condition `[cond` i`]` enclosed in its formalization. The checker for actual applicability simply compares whether the linker indicated in the input compound by SMOR is one of the licensed `[link 1, ..., link` m`]`. For the L2P constraints, the programmatic checkers are swapped as well, following the formal tests. Since we could only access morphological, derivational, and phonological properties of single constituents in CELEX-nouns, constraints involving semantic or lexical information of the first/second constituent as well as the semantic relation between the two were excluded from automated processing. In total, pairs of checkers for 40 constraints were implemented.

We ran all available checkers against every compound in GeCoDB_v06. As a result, we automatically discovered and marked all compounds to which the corresponding constraints are potentially and actually applicable (individually). Finally, we employed the mathematical formalizations from section 4.2 to calculate the two variants of coverage and regularity for each constraint.

We find it crucial to emphasize the role of the structured linguistic data for our study. The strict triplet structure of SMOR analyses in GeCoDB_v06 allowed to establish a dense connection to the CELEX-nouns database through reliable lookup of compound constituents in CELEX-nouns. In its turn, the highly structured tabular format of CELEX-nouns played a principal role in stable and predictable programmatic access to the desired linguistic properties of the constituents. It is therefore the structural concert of the two data collections that were crucial for reliable automated testing for empirical constraint application in the course of our corpus study.

5. Results & Discussion

In total, we obtained item and type coverage and regularity for 39 constraints[22] (23 P2L, 15 L2P, and the default constraint). Figure 1 demonstrates the distribution of the obtained values. The entire constraint space with the mapping is depicted illustratively in Appendix D, and the precise item and type

[22] One of the 40 constraints selected for automated processing did not cover any compounds in the dataset (namely, (26), see Appendix E).

coverage and regularity values for each analyzed constraint are provided in Appendix E.

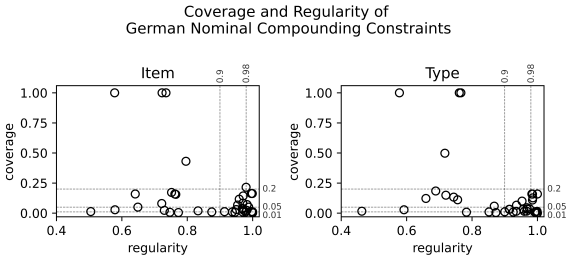

Figure 1: Distribution of the obtained coverage and regularity values. Each constraint appears twice: on the left and on the right.

Both coverage and regularity values fluctuate significantly. Item and type coverage stretch over the entire range of values: they effectively approximate 0.0 for a few constraints and reach 1.0 for three constraints. The standard deviation (SD) of 0.26 and 0.27 for the two variants, respectively, further confirms a large spread of values. Nevertheless, it can be observed that most of the coverage values concentrate at the bottom part of the plot. Indeed, 34 out of 39 constraints lay below the threshold of 0.2 for both item and type coverage. Going further, combined coverage falls under 0.05 for almost half of the inspected constraints (19 in total). Three constraints cover very restricted groups of nouns with combined coverage amounting to < 0.01.

Regularity, on the contrary, is distributed more compactly. All item and type regularity values lie in a range $(0.46; 1.0]$. They are also spread more evenly than the coverage values, with significantly lower SD of 0.14 for both variants. Even though the full range of values is densely occupied with no significant gaps, there is still a visible shift of values towards the right edge of the axis. Thus, slightly over half of the analyzed constraints (20 in total) reach a combined regularity of over 0.9. However, only six of them appear to be approximately regular with a combined regularity of > 0.98.

We observe that moderate and low values of coverage and regularity of both types are distributed relatively evenly between the P2L and the L2P constraints, whereas high values are distributed in a nearly complementary manner. Specifically, the entire range $(0.02; 1.0]$ of combined coverage values is occupied by three P2L constraints plus the default constraint. On the other hand, the upper range $(0.98; 1.0]$ of combined regularity is populated almost exclusively by the values of L2P constraints, with five out of six constraints in this range belonging to this category. These observations suggest that the restrictions on the noun classes that may attach certain linkers gravitate towards reliable application, while the patterns of linker choice prioritize broader predictive scope. The plot of applicability statistics of the P2L vs. the L2P constraints is provided in Appendix C.

A productivity bias as exemplified by constraint (22) is detected for six constraints in total. Type regularity of constraints (4), (22), and (42) is at least 10% lower than their item productivity. Similarly to (22), the bias in (4) and (42) is produced by smaller sets of items that productively form compounds that violate these constraints. Type regularity of constraints (19), (25), and (44) is, on the contrary, higher than their item regularity by at least 10%. This points to an opposite situation: the bias comes from sizable sets of items that produce small groups of compounds that tend to violate these constraints. For instance, 140 out of 1517 covered compounds that violate (19) are formed from 8 out of 27, or almost a third of unique first constituents.

No correlation between coverage and regularity is observed, neither for the item nor for the type variant. We conclude that these two properties are independent.

To assess the extent to which the collected constraints are supported by the empirical data, we assigned a threshold of minimal item regularity for each unique quantifier found in the constraints definitions ("mostly", "regularly", etc.). Item regularity was chosen as a more balanced metric as compared to type regularity. As a result, the scale demonstrated in figure 2 was established[23]. A constraint was considered *supported* if its item regularity equaled or exceeded the threshold corresponding to its textual quantifier (e.g., "usually" vs. item regularity of ≥ 0.9). A constraint was considered *partially supported* if its item regularity equaled or exceeded the threshold of the previous level on the scale but not its own (e.g., "usually" vs. item regularity in range $[0.6; 0.9)$). A constraint was regarded as *not supported* if its item regularity was below the threshold of the previous level (e.g., "usually" vs. item regularity of < 0.6).

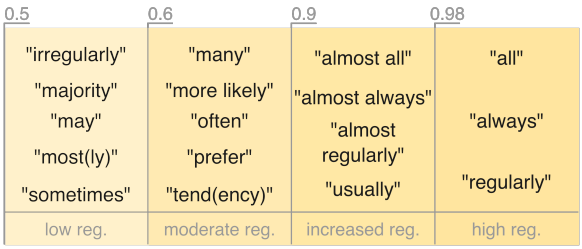

Figure 2: Quantifiers and the corresponding item regularity thresholds by regularity (reg.) levels.

Most of our findings corroborate the qualitative evidence from previous literature. Specifically, 28 out of 39 analyzed constraints are supported by the

[23] Please note that the threshold for the regular level is reduced to 0.97 for constraint whose item coverage equals or falls under 0.01. The reduction counterbalances an increased sensitivity of item regularity to counterexamples resulting from a very small number of items.

corpus data as defined in the previous paragraph. In the next paragraphs, we inspect the constraints that are partially/not supported by the empiric data on a number of illustrative examples.

Seven constraints marked as highly regular in the literature and one constraint marked as moderately regular did not confirm their status on massive data but appeared to be partially supported. These are P2L constraints (3), (29), (31) and L2P constraints (47), (48), (52), (56), (61). For instance, constraint (3) which is formulated as "First constituents that are constituted by nouns that end in a full vowel always adopt *-ø-*" (Kürschner, 110; Schäfer and Pankratz, 2018, 9) is consistently violated by compounds built with foreign nouns ending with a stressed *-ee*/*-ie* (e.g., Idee**n**buch '*book of ideas*', Kalorie**n**gehalt '*calorie content*', Melodie**n**folge '*melody sequence*'). Otherwise, rare counterexamples with native words are found, such as Schuh**e**kauf '*purchase of shoes*'. The other seven constraints from this group likewise have numerous counterexamples.

Constraints (19), (25), (33) (all P2L) were not supported by the empirical data. The first disproved constraint (19) is formulated in Ortner and Müller-Bollhagen (1991, 83, 107): "First constituents that are constituted by derived nouns with suffixes *-bold*, *-nis*, *-rich*, *-at*, *-al* that build the plural form with *-e* attach *-ø-* regularly." With over 190 counterexamples of compounds with first constituents ending with *-at* and attaching *-s-* (e.g., Dekanat**s**veranstaltung '*deanery event*', Referat**s**vorbereitung 'presentation preparation'), we record item regularity of 0.747 and type regularity 0.873 for this constraint. For the covered compounds with the rest of the suffixes (excluding *-at*), the constraint is regular, indeed. The second not supported constraint (25) is formulated as "First constituents that are constituted by deverbal nouns that end in suffix *-en* attach *-s-* regularly." (Eisenberg, 2013, 230; Nübling and Szczepaniak, 2013, 78). The counterexamples include two groups of compounds. The first group consists of 104 *ø*-linked variants of existing compounds conforming to (25). In most of these pairs, the *ø*-linked variant has a significantly lower frequency (e.g., Leben**s**stil vs. Leben*ø*stil '*lifestyle*' with DeReKo word counts 63234 and 293, respectively). This observation might signalize that constraint (25) may be evolving towards lower restrictiveness in contemporary German. The second group includes 82 further compounds with *-ø-* formed from 13 first constituents (Beben*ø*gebiet '*quake zone*', Braten*ø*soße '*gravy*', Husten*ø*saft '*cough syrup*', etc.). Overall, the constraint reaches only 0.773 of item regularity (but 0.926 of type regularity), which corresponds to the moderate regularity level. Finally, constraint (33) "First constituents that are constituted by feminine nouns that end in schwa adopt *-(e)n-* regularly." (Eisenberg, 2013, 229; Krott et al., 2007, 3) is not supported by the data. Even though Fuhrhop and Kürschner (2015, 573), Kürschner, 112, Ortner and Müller-Bollhagen (1991, 93) point out that *-ø-* is also probable after such nouns if the schwa is a deadjectival or a deverbal suffix, it seems that the real number of such cases in language has been underestimated in the literature. We detected over 5.3k compounds with 109 various deverbal and deadjectival feminine first constituents that do not insert an explicit link (e.g., Reise*ø*ziel '*destination*', Glätte*ø*stelle '*slippery spot*'). Even apart from the derived first constituents, a very large group of counterexamples formed from 203 non-derived first constituents (Analyse*ø*bericht '*analysis report*', Bronze*ø*lack '*bronze lacquer*', Ehe*ø*mann '*husband*', Messe*ø*besucher '*visitor of a fair*', and much more) adds to this significant number, resulting in over 9.6k compounds violating the constraint. In total, this constraint achieves an item regularity of only 0.765 and a type regularity of 0.744, which, again, belongs to the moderate regularity level.

The provided examples reveal that for some constraints, not the regularity level but the conditions on potential applicability seem to be formulated incorrectly. Consider constraint (19). As already mentioned, it applies regularly to compounds with first constituents ending in all suffixes specified in its definition, besides *-at*. Correspondingly, after removing this suffix from the definition, (19) would be fully supported by the empirical data. In Appendix F, we provide a list of our suggestions on how to correct the definitions of the constraints that are supported only partially or not supported. We additionally implemented dedicated Python checkers for the corrected versions of constraint as proposed by us, and we report that all corrected versions are supported by the corpus data.

6. Conclusion

The linker choice in German N+N compounds is highly unpredictable. It is governed by a set of (irregular) constraints whose conditions on potential applicability are defined over various linguistic domains. The present study collects constraints documented in the linguistics literature and empirically investigates their application over a massive compounds corpus. The study verified 39 constraints collected in the relevant literature. It confirmed 28 constraints, found smaller inconsistencies towards the corpus data in eight constraint definitions, and identified major discrepancies towards the data for three constraints. In sum, it proposes novel findings on German nominal compounding and calls attention to the necessity to support qualitative linguistic studies by empirical estimations.

7. Limitations

In future research, we plan to implement the missing semantic and lexical constraints and obtain a full picture of nominal constraints' coverage and regularity. An open question is whether the type of linguistic information that a constraint employs has any relation to the constraint's coverage and regularity. Finally, during the experiment we detected a large number of constraint conflicts over various compounds. We believe that investigating constraint conflict resolution is a crucial research direction for understanding the complex interaction of the constraints in the process of linker choice. This research question is therefore to be addressed in future studies.

8. Acknowledgments

I would like to thank my PhD advisor, Erhard Hinrichs, for extensive comments on the paper, insightful discussions, and guidance throughout the study. His academic support was crucial to the success of this project. I am grateful to the three anonymous reviewers for the helpful comments. I would like to acknowledge the contribution of Marie Hinrichs, who did the final proofreading. Any remaining errors are solely my responsibility.

9. Bibliographical References

Gerhard Augst. 1975. *Untersuchungen zum Morpheminventar der deutschen Gegenwartssprache*. Narr, Tübingen.

Marco Baroni, Johannes Matiasek, and Harald Trost. 2002. Wordform- and Class-based Prediction of the Components of German Nominal Compounds in an AAC System. In *COLING 2002: The 19th International Conference on Computational Linguistics*.

Peter Eisenberg. 2013. *Grundriss der deutschen Grammatik*, 4 edition. J.B. Metzler Stuttgart, Stuttgart. EBook published 27 August 2016.

Nanna Fuhrhop and Sebastian Kürschner. 2015. *32. Linking elements in Germanic*, pages 568–582. De Gruyter Mouton, Berlin, München, Boston.

Maria Koliopoulou. 2014. How close to syntax are compounds? Evidence from the linking element in German and Modern Greek compounds. *Rivista di Linguistica*, 26(2):51–70.

Kristin Kopf. 2017. *Fugenelement und Bindestrich in der Compositions-Fuge*, pages 177–204. De Gruyter, Berlin, Boston.

Andrea Krott, Gary Libben, Gonia Jarema, Wolfgang Dressler, Robert Schreuder, and Harald Baayen. 2004. Probability in the Grammar of German and Dutch: Interfixation in Triconstituent Compounds. *Language and Speech*, 47(1):83–106. PMID: 15298331.

Andrea Krott, Robert Schreuder, R. Harald Baayen, and Wolfgang U. Dressler. 2007. Analogical effects on linking elements in German compound words. *Language and Cognitive Processes*, 22(1):25–57.

Sebastian Kürschner. Verfugung-s-nutzung kontrastiv. Zur Funktion der Fugenelemente im Deutschen und Dänischen. *Tijdschrift voor Skandinavistiek*, 26(2).

Sebastian Kürschner. 2010. Fuge-n-kitt, voeg-en-mes, fuge-masse und fog-e-ord - Fugenelemente im Deutschen, Niederländischen, Schwedischen und Dänischen : ein Grenzfall der Morphologie im Sprachkontrast. (Linking elements in German, Dutch, Swedish, and Danish contrast). In *Dammel, Antje; Kürschner, Sebastian; Nübling, Damaris (Hrsg.): Kontrastive Germanistische Linguistik*, volume 206-209 of *Germanistische Linguistik*, pages 827–862. Olms, Hildesheim.

Gary Libben, Monika Boniecki, Marlies Martha, Karin Mittermann, Katharina Korecky-Kröll, and Wolfgang U. Dressler. 2009. Interfixation in German compounds: What factors govern acceptability judgements? *Rivista di Linguistica*, 21(1):149–180.

Martin Neef and Susanne R. Borgwaldt. 2012. *Fugenelemente in neugebildeten Nominalkomposita*, pages 27–56. De Gruyter, Berlin, Boston.

Damaris Nübling and Renata Szczepaniak. 2008. On the way from morphology to phonology: German linking elements and the role of the phonological word. *Morphology*, 18(1):1–25.

Damaris Nübling and Renata Szczepaniak. 2013. Linking elements in German Origin, Change, Functionalization. *Morphology*, 23(1):67–89.

Lorelies Ortner and Elgin Müller-Bollhagen. 1991. *Hauptteil 4 Substantivkomposita*. De Gruyter, Berlin, Boston.

Roland Schäfer. 2018. *Einführung in die grammatische Beschreibung des Deutschen*. Number 2 in Textbooks in Language Sciences. Language Science Press, Berlin.

Roland Schäfer and Elizabeth Pankratz. 2018. The plural interpretability of German linking elements. *Morphology*, 28(4):325–358.

Carmen Scherer. 2012. *Vom Reisezentrum zum Reise Zentrum. Variation in der Schreibung von N+N-Komposita*, pages 57–82. De Gruyter, Berlin, Boston.

Barbara Schlücker. 2023. Compounding and Linking Elements in Germanic.

Barbara Schlücker. 2012. *Die deutsche Kompositionsfreudigkeit. Übersicht und Einführung*, pages 1–26. De Gruyter, Berlin, Boston.

Helmut Schmid, Arne Fitschen, and Ulrich Heid. 2004. SMOR: A German Computational Morphology Covering Derivation, Composition, and Inflection. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).

Heide Wegener. 2003. Entstehung und Funktion der Fugenelemente im Deutschen, oder: warum wir keine Autosbahn haben. *Linguistische Berichte*, 196:425–457.

Heide Wegener. 2005. Das Hühnerei vor der Hundehütte : von der Notwendigkeit historischen Wissens in der Grammatikographie des Deutschen.

10. Language Resource References

Baayen, R. H. and Piepenbrock, R. and Gulikers, L. 1995. *CELEX2*. Linguistic Data Consortium, ISLRN 204-698-863-053-1. Web Download.

IDS. 2026. *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2026-I*. Leibniz-Institut für Deutsche Sprache. PID https://www.ids-mannheim.de/dereko. Release vom 19.01.2026.

Schäfer, Roland and Pankratz, Elizabeth. 2018. *Dataset: The plural interpretability of German linking elements ("Morphology")*. Zenodo. PID https://doi.org/10.5281/zenodo.1323211.

A. Sources of Constraint Definitions

Source	Addressed Constraints
Eisenberg (2013, chapter 6.2)	(1), (3), (6), (7), (8), (10), (11), (23), (25), (27), (33), (34), (35), (55), (59), (60)
Fuhrhop and Kürschner (2015)	(1), (3), (23), (33), (40), (41), (42)
Koliopoulou (2014)	(23), (27), (36)
Kopf (2017)	(1), (27), (40), (48)
Krott et al. (2007)	(1), (16), (23), (33), (34), (64)
Kürschner	(1), (23), (25), (27), (29), (40), (42), (48)
Kürschner (2010)	(3), (23), (27), (28), (32), (40), (43), (44), (45), (47), (48), (51), (52), (53), (55), (56), (60), (61)
Libben et al. (2009)	(1), (33)
Neef and Borgwaldt (2012)	(36)
Nübling and Szczepaniak (2008)	(1), (6), (23), (27), (35), (38), (39), (41), (42), (48), (51), (55), (59), (60), (63)
Nübling and Szczepaniak (2013)	(3), (6), (9), (10), (13), (16), (17), (23), (25), (27), (31), (34), (35), (37), (41), (43), (47), (48), (49), (51), (55), (60), (62)
Ortner and Müller-Bollhagen (1991, chapter A.5)	(1), (2), (3), (4), (5), (7), (8), (10), (11), (12), (14), (16), (17), (18), (19), (23), (24), (25), (26), (27), (28), (30), (31), (32), (33), (34), (35), (36), (38), (39), (46), (49), (50), (51), (53), (54), (55), (57), (58), (62)
(Schäfer, 2018, chapter 8.1)	(3), (16), (23), (35), (64)
Schäfer and Pankratz (2018)	(1), (5), (10), (13), (15), (16), (23), (29), (33), (42), (43), (60), (62)
Schlücker (2012)	(1)
Schlücker (2023)	(23), (38), (64)
Wegener (2003)	(23), (42)
Wegener (2005)	(28), (38), (42), (47), (64)

Table 2: Linguistics literature providing the evidence for the nominal compounding constraints in German.

B. Calculation of Applicability Statistics: Example on Constraint (22)

Constraint (22) is potentially applicable to 2798 compounds formed from 32 unique first constituents. GeCoDB_v06 contains 282768 unique compounds built with 3613 unique first constituents. Then, type coverage of cov_{type} of (22) equals

$$\text{cov}_{type} = \frac{2798}{282768} \approx 0.01,$$

and item coverage cov_{item} of (22)

$$\text{cov}_{item} = \frac{32}{3613} \approx 0.009.$$

Constraint (22) is actually applicable to 2384 compounds. Therefore, type regularity reg_{type} of (22) is calculated as

$$\text{reg}_{type} = \frac{2384}{2798} \approx 0.852.$$

Thirty-one out of 32 first constituents always build compounds that conform to (22), so their group regularity equals 1.0. The remaining first constituent Liebe '*love*' builds 434 compounds in total, from which only 20 conform to (22), estimating Liebe's group regularity at $\frac{20}{434} \approx 0.046$. Then, item regularity reg_{item} of (22) amounts to

$$\text{reg}_{item} = \frac{31 \times 1.0 + 0.046}{32} \approx 0.97.$$

C. Coverage and Regularity Values: P2L vs. L2P

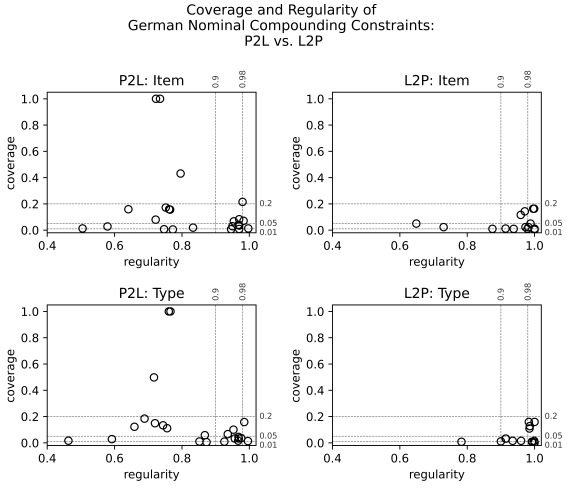

Figure 3: Distribution of the obtained coverage and regularity values with respect to the P2l/L2P type. Each P2l/L2P constraint appears twice: on the top and on the bottom.

D. Constraint Space Plot

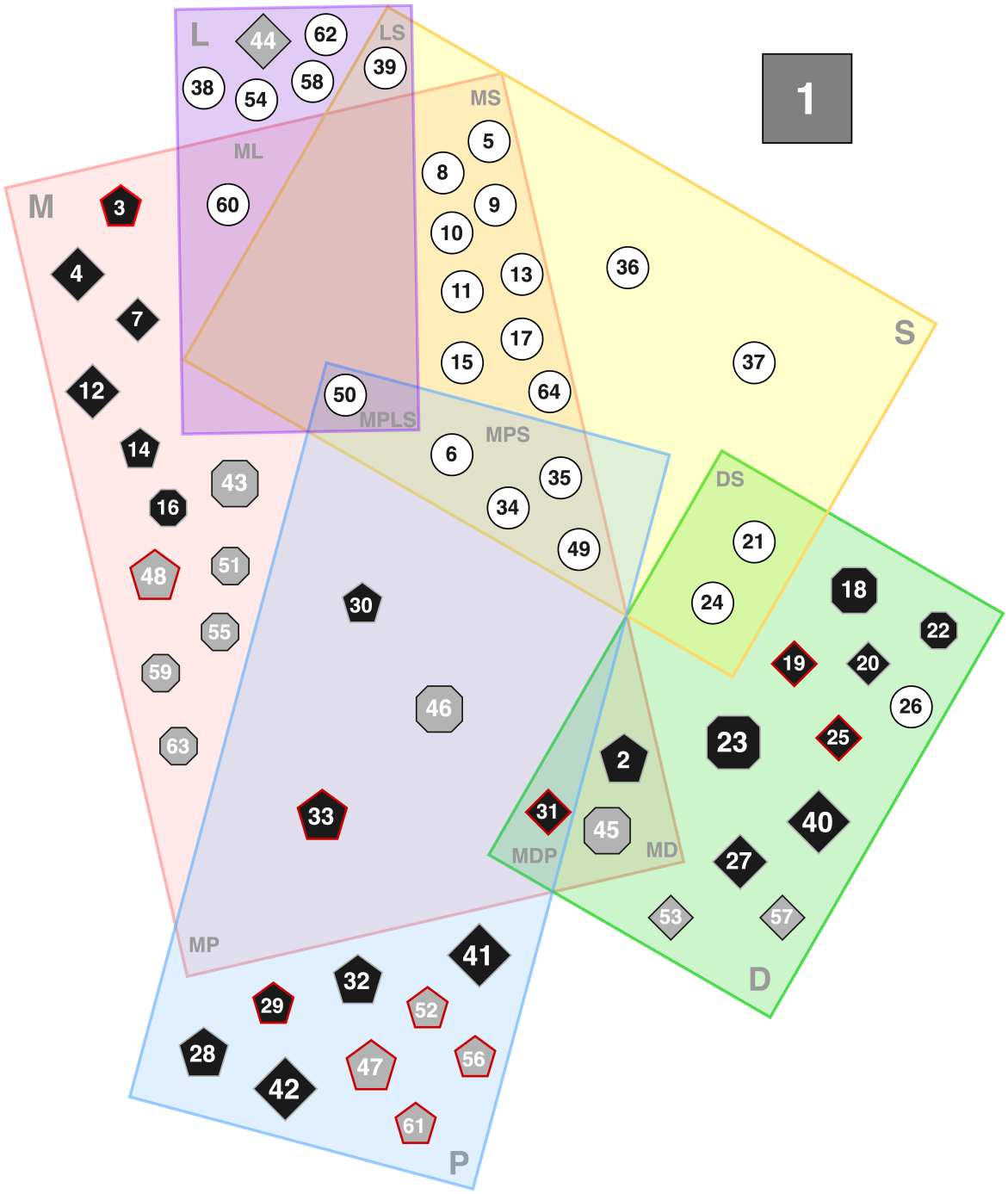

Figure 4: Overall illustration of the constraint space. Constraint (1) is the default constraint; it is placed separately outside of the box. For the other constraints, the size of the shape indicates the magnitude of its item coverage: small shapes — < 0.05, mid-sized shapes — in range $(0.05; 0.2)$, big shapes — ≥ 0.2. The number of angles in the shape is associated with the magnitude of item regularity: rhombs designate constraints with low and moderate regularity level, pentagons — increased regularity level, and octagons — high regularity level. Black shapes indicate P2L, and grey — L2P constraints. A red border signalizes that the constraint is partially/not supported by the empirical data. White circles represent constraints that could not be processed automatically. Large rectangles of different colors designate groups of constraints by the type of linguistic information they exploit: red is the morphological type, green is derivational, blue — phonological, yellow — semantic, and purple represents the lexical type. Mixed type groups are situated on the overlapping areas of the corresponding types. Each non-empty area is supplied with an abbreviation of the (mixed) type it represents. For the corresponding constraint definitions and the precise statistics values see Appendix E.

E. Constraint Definitions & Properties

Below is the list of German nominal compounding constraints summarized from the relevant literature. Each entry starts with the title and the definition of the corresponding constraint. The quantifiers are underlined. The grey tags correspond to the P2L/L2P type of the constraint, type of the constraint by employed linguistic information (abbreviated according to figure D), and the target item of the constraint (except for the default constraint (1)). For the 39 constraints selected for automated processing, item and type coverage and regularity are provided in form of a table, and the support level is specified under it. Examples, counterexamples (translations omitted), and the relevant evidence from the linguistics literature concludes the entry. The entries are separated by horizontal bars.

Constraint (1)

The majority of compounds have a zero linker. Zero linker is default for German compounds.

default | first + second constituent

	item	type
coverage	1.0	1.0
regularity	0.578	0.578

Support level: *supported*

Examples: Landøkreis, Hausøtür, Stadtømauer
Counterexamples: Land**es**regiesrung, H**äu**s**er**block, St**ä**dt**e**tag

Sources: Eisenberg (2013, 226), Fuhrhop and Kürschner (2015, 569), Kopf (2017, 177), Krott et al. (2007, 3), Kürschner, 107, Libben et al. (2009, 150, 156), Nübling and Szczepaniak (2008, 2), Ortner and Müller-Bollhagen (1991, 50, 52, 54, 103), Schäfer and Pankratz (2018, 8, 9), Schlücker (2012, 9)

Constraint (2)

First constituents that are constituted by simplex masculine or neuter nouns that build the plural form with a zero ending and simplex or complex feminine nouns that build the plural form with a zero ending attach -ø- almost regularly.

P2L | MD | first constituent

	item	type
coverage	0.067	0.066
regularity	0.954	0.936

Support level: *supported*

Examples: Schlittenøfahrt, Wagenørad, Uferøpromenade
Counterexamples: Orden**s**bruder, Teufel**s**werk, Frieden**s**angebot

Sources: Ortner and Müller-Bollhagen (1991, 106)

Constraint (3)

First constituents that are constituted by nouns that build the plural form with *-s* attach -ø- regularly.

P2L | M | first constituent

	item	type
coverage	0.035	0.033
regularity	0.97	0.957

Support level: *partially supported*

Examples: Balkonømöbel, Hoteløkette, Bobøbahn
Counterexamples: Uhu**s**nest, Deck**s**balken, Interim**s**phase, Karneval**s**kostüm

Sources: Eisenberg (2013, 227), Fuhrhop and Kürschner (2015, 572), Kürschner (2010, 9), Nübling and Szczepaniak (2013, 78), Ortner and Müller-Bollhagen (1991, 83, 106), Schäfer (2018, 229)

Constraint (4)

First constituents that are constituted by nouns that build the plural form with *-e* mostly have -ø-.

P2L | M | first constituent

	item	type
coverage	0.159	0.184
regularity	0.762	0.689

Support level: *supported*

Examples: Tischødecke, Projektøgruppe, Pilzøsammler
Counterexamples: Tag**e**gericht, Projekt**e**macherei, Hund**e**leine

Sources: Ortner and Müller-Bollhagen (1991, 83)

Constraint (5)

First constituents that are constituted by nouns that build the plural form with *-e* attach *-e-* irregularly in compounds in which the second constituent forces a plural meaning of the given first constituent.

P2L | MS | first constituent

Examples: Projekt**e**macherei, Tag**e**buch, Punkt**e**stand
Counterexamples: Pilzøsammler, Brotøkorb, Jahrøzehnt, Tag**e**lohn

Sources: Ortner and Müller-Bollhagen (1991, 103), Schäfer and Pankratz (2018, 29)

Constraint (6)

First constituents that are constituted by monosyllabic animal designations that build the plural form with *-e* attach *-e-* irregularly.

P2L MPS first constituent

Examples: Hund**e**leine, Swein**e**stall, Pferd**e**wagen
Counterexamples: Hund**s**stern, Schweinøkram, Schafføstall

Sources: Eisenberg (2013, 228), Nübling and Szczepaniak (2008, 11), Nübling and Szczepaniak (2013, 81)

Constraint (7)

First constituents that are constituted by nouns that build the plural form with *-(")er* mostly attach *-ø-* or *-(")er-*.

P2L M first constituent

	item	type
coverage	0.019	0.058
regularity	0.833	0.868

Support level: *supported*

Examples: Buchødeckel, B**ü**ch**er**regal, Kind**er**jacke
Counterexamples: Kind**es**alter, Mann**s**volk, Rindøfleisch

Sources: Eisenberg (2013, 229), Ortner and Müller-Bollhagen (1991, 99)

Constraint (8)

First constituents that are constituted by nouns that build the plural form with *-(")er* attach *-ø-* in a majority of compounds in which the second constituent forces a singular meaning of the given first constituent.

P2L MS first constituent

Examples: Buchødeckel, Grabøstein, Blattøfläche
Counterexamples: M**ä**nn**er**herz, Kind**er**bild, Buchøhandel

Sources: Eisenberg (2013, 229), Ortner and Müller-Bollhagen (1991, 99, 108)

Constraint (9)

First constituents that are constituted by nouns that build the plural form with *-(")er* tend to attach *-ø-* in compounds in which the second constituent forces a mass meaning of the given first constituent.

P2L MS first constituent

Examples: Glasøauge, Krautøsalat, Hornøhaut
Counterexamples: Krautøgarten

Sources: Nübling and Szczepaniak (2013, 81)

Constraint (10)

First constituents that are constituted by nouns that build the plural form with *-(")er* attach *-(")er-* more likely in compounds in which the second constituent forces a plural meaning of the given first constituent.

P2L MS first constituent

Examples: B**ü**ch**er**regal, Gr**ä**b**er**feld, Kr**ä**ut**er**frau
Counterexamples: Buchøhandel, Bildøband, Kind**er**jacke, M**ä**nn**er**herz

Sources: Eisenberg (2013, 229), Nübling and Szczepaniak (2013, 80, 81), Ortner and Müller-Bollhagen (1991, 99), Schäfer and Pankratz (2018, 29)

Constraint (11)

First constituents that are constituted by person and animal designations that build the plural form with *-(")er* may attach *-(")er-* regardless of the singular/-plural interpretation of the given first constituent within the compound.

P2L MS first constituent

Examples: Kind**er**jacke, H**ü**hn**er**ei, Rind**er**brust

Sources: Eisenberg (2013, 229), Ortner and Müller-Bollhagen (1991, 98)

Constraint (12)

First constituents that are constituted by nouns that build the plural form with *-e* and umlaut mostly have *-ø-*.

P2L M first constituent

	item	type
coverage	0.08	0.111
regularity	0.722	0.756

Support level: *supported*

Examples: Handøfläche, Fruchtøjoghurt, Arztøpraxis
Counterexamples: H**ä**nd**e**druck, **Ä**rzt**e**streik, Fr**ü**cht**e**tee

Sources: Ortner and Müller-Bollhagen (1991, 83, 107)

Constraint (13)

First constituents that are constituted by nouns that build the plural form with *-e* and umlaut attach *-"e-* irregularly in compounds in which the second constituent forces a plural meaning of the given first constituent.

P2L MS first constituent

Examples: H**ä**nd**e**druck, **Ä**rzt**e**streik, G**ä**st**e**buch
Counterexamples: Baumøgruppe, Zugøverkehr

Sources: Nübling and Szczepaniak (2013, 81), Schäfer and Pankratz (2018, 29)

Constraint (14)

First constituents that are constituted by nouns that build the plural form with a zero ending and umlaut almost always have *-ø-*.

P2L | M | first constituent

	item	type
coverage	0.007	0.016
regularity	0.946	0.967

Support level: *supported*

Examples: Apfeløbaum, Klosterøkirche, Mutterøsprache
Counterexamples: Br**ü**derøgemeinde, M**ü**tterøzentrum, V**ä**terøaufbruch

Sources: Ortner and Müller-Bollhagen (1991, 83, 106)

Constraint (15)

First constituents that are constituted by nouns that build the plural form with a zero ending and umlaut attach *-ø-* with umlaut irregularly in compounds in which the second constituent forces a plural meaning of the given first constituent.

P2L | MS | first constituent

Examples: Br**ü**derøgemeinde, M**ü**tterøzentrum, V**ä**terøaufbruch
Counterexamples: Vogeløfutter

Sources: Schäfer and Pankratz (2018, 29)

Constraint (16)

First constituents that are constituted by mixed masculine and neuter nouns almost always attach *-s-*, *-ø-*, or *-(e)n-*.

P2L | M | first constituent

Examples: Staat**s**amt, Staat**en**bund, Bettøanzug, Auge**n**lied
Counterexamples: Schmerz**ens**schrei

Sources: Krott et al. (2007, 3), Nübling and Szczepaniak (2013, 80), Ortner and Müller-Bollhagen (1991, 92), Schäfer (2018, 229), Schäfer and Pankratz (2018, 9)

Constraint (17)

First constituents that are constituted by mixed masculine and neuter nouns attach *-(e)n-* more likely in compounds in which the second constituent forces a plural meaning of the given first constituent.

P2L | MS | first constituent

Examples: Staat**en**bund, Bett**en**zahl, Strahl**en**belastung
Counterexamples: Motor**en**geräusch, Professor**en**titel, Auge**n**lied, Interesse**n**bereich

Sources: Nübling and Szczepaniak (2013, 80), Ortner and Müller-Bollhagen (1991, 92)

Constraint (18)

First constituents that are constituted by derived nouns with suffixes *-er*, *-ler*, *-ner*, *-el*, *-sel*, *-chen*, *-lein* that build the plural form with a zero ending attach a zero linking element regularly.

P2L | D | first constituent

	item	type
coverage	0.071	0.045
regularity	0.983	0.968

Support level: *supported*

Examples: Reiterøschwert, Gürteløschnalle, Brötchenøgeber
Counterexamples: Alter**s**abstand, Reiter**s**mann

Sources: Ortner and Müller-Bollhagen (1991, 56, 82, 106)

Constraint (19)

First constituents that are constituted by derived nouns with suffixes *-bold*, *-nis*, *-rich*, *-at*, *-al* that build the plural form with *-e* attach *-ø-* regularly.

P2L | D | first constituent

	item	type
coverage	0.007	0.005
regularity	0.747	0.873

Support level: *not supported*

Examples: Erlaubnisøschein, Formatøvorlage, Personaløausweis
Counterexamples: Radikal**e**fänger, Internat**s**leitung, Magistrat**s**verfassung

Sources: Ortner and Müller-Bollhagen (1991, 83, 107)

Constraint (20)

First constituents that are constituted by deverbal feminine nouns with a schwa suffix mostly attach *-ø-*.

P2L | D | first constituent

	item	type
coverage	0.028	0.028
regularity	0.579	0.592

Support level: *supported*

Examples: Abgabeøsoll, Sorgeøpflicht, Lageøplan
Counterexamples: Abgabe**n**ordnung, Anzeige**n**blatt, Lüge**n**detektor

Sources: Ortner and Müller-Bollhagen (1991, 93)

Constraint (21)

First constituents that are constituted by deverbal feminine nouns with a schwa suffix tend to attach *-(e)n-* in compounds in which the second constituent forces a plural reading of the given first constituent.

P2L | DS | first constituent

Examples: Abgabe**n**ordnung, Anzeige**n**blatt, Lüge**n**detektor
Counterexamples: Einnahme**ø**buch, Anzeige**ø**tafel

Sources: Ortner and Müller-Bollhagen (1991, 93)

Constraint (22)

First constituents that are constituted by deadjectival feminine nouns with a schwa suffix almost always attach *-(e)n-* or *-ø-*. Each of the two linkers is preferred in about the same number of cases.

P2L | D | first constituent

	item	type
coverage	0.009	0.01
regularity	0.97	0.852

Support level: *supported*

Examples: Fläche**ø**maß, Hitze**ø**welle, Tiefe**n**meter, Größe**n**klasse
Counterexamples: Liebe**s**brief

Sources: Fuhrhop and Kürschner (2015, 573), Kürschner, 112, Ortner and Müller-Bollhagen (1991, 93)

Constraint (23)

First constituents that are constituted by derived nouns with suffixes *-(ig)keit*, *-heit*, *-schaft*, *-ung*, *-sal*, *-ing*, *-ling*, *-tum*, *-um*, *-ion*, also *-ität* and its allomorphs attach *-s-* regularly.

P2L | D | first constituent

	item	type
coverage	0.215	0.158
regularity	0.98	0.985

Support level: *supported*

Examples: Gesundheit**s**amt, Gesellschaft**s**politik, Schicksal**s**drama, Rarität**s**wert, Eigentum**s**recht
Counterexamples: Minderheit**en**recht, Kuriosität**en**museum, Nationalität**en**kampf

Sources: Eisenberg (2013, 230), Fuhrhop and Kürschner (2015, 572), Koliopoulou (2014, 61), Krott et al. (2007, 3), Kürschner, 107, 115, Kürschner (2010, 20, 21), Nübling and Szczepaniak (2008, 10, 20, 23), Nübling and Szczepaniak (2013, 77), Ortner and Müller-Bollhagen (1991, 73, 83, 88, 89, 94), Schäfer (2018, 229), Schäfer and Pankratz (2018, 8), Schlücker (2023, 19), Wegener (2003, 448)

Constraint (24)

First constituents that are constituted by derived nouns with suffix *-ität* and its allomorphs prefer *-(e)n-* in compounds in which the second constituent forces a plural meaning of the given first constituent.

P2L | DS | first constituent

Examples: Rarität**en**sammlung, Kuriosität**en**händler, Aktualitäten**en**kino

Sources: Ortner and Müller-Bollhagen (1991, 94)

Constraint (25)

First constituents that are constituted by deverbal nouns that end in suffix *-en* attach *-s-* regularly.

P2L | D | first constituent

	item	type
coverage	0.005	0.009
regularity	0.773	0.926

Support level: *not supported*

Examples: Leben**s**mittel, Essen**s**gewohnheit, Unternehmen**s**bereich
Counterexamples: Essen**ø**portion, Leben**ø**geschäft, Husten**ø**tropfen, Beben**ø**gebiet

Sources: Eisenberg (2013, 230), Kürschner, 107, Nübling and Szczepaniak (2013, 78), Ortner and Müller-Bollhagen (1991, 89)

Constraint (26)

First constituents that are constituted by derived feminine nouns with suffix *-in* always attach *-(e)n-*. (The suffix -in adjusts orthographically in this case and becomes -inn.)

P2L | D | first constituent

Examples: Lehrerinn**en**mentalität, Freundinn**en**gruppe, Schülerinn**en**rat
Counterexamples: Königin**ø**mutter

Sources: Ortner and Müller-Bollhagen (1991, 94)

Constraint (27)

First constituents that are constituted by prefixed deverbal nouns exhibit a tendency to attach *-s-*.

P2L | D | first constituent

	item	type
coverage	0.159	0.122
regularity	0.641	0.659

Support level: *supported*

Examples: Anspruch**s**haltung, Eintrag**s**frist, Bedarf**s**fall, Verfall**s**datum

Counterexamples: Anruf**ø**beantworter, Überfall**ø**kommando, Bestand**ø**teil

Sources: Eisenberg (2013, 230), Koliopoulou (2014, 61), Kürschner, 107, Kürschner (2010, 21), Kopf (2017, 190), Nübling and Szczepaniak (2008, 18, 19), Nübling and Szczepaniak (2013, 78), Ortner and Müller-Bollhagen (1991, 89)

Constraint (28)

First constituents that are constituted by nouns that end in a sibilant or in a consonant cluster including [s] mostly adopt *-ø-*.

P2L | P | first constituent

	item	type
coverage	0.083	0.1
regularity	0.97	0.953

Support level: *supported*

Examples: Gefäß**ø**system, Fisch**ø**öl, Notiz**ø**buch, Herbst**ø**anfang
Counterexamples: G**ä**st**e**buch, Geist**er**haus

Sources: Kürschner (2010, 19), Ortner and Müller-Bollhagen (1991, 89, 109), Wegener (2005, 181)

Constraint (29)

First constituents that are constituted by nouns that end in a full vowel always adopt *-ø-*.

P2L | P | first constituent

	item	type
coverage	0.041	0.036
regularity	0.968	0.976

Support level: *partially supported*

Examples: Auto**ø**bahn, Uhu**ø**paar, Kuh**ø**milch, Gummi**ø**bär
Counterexamples: Uhu**s**nest, Idee**n**liste, Kalorie**n**wert, Schuh**e**kauf

Sources: Kürschner, 110, Schäfer and Pankratz (2018, 9)

Constraint (30)

First constituents that are constituted by feminine nouns that end with stressed *-ei*, *-ie*, *-ur*, also stressed or unstressed *-ik* usually attach *-ø-*.

P2L | MP | first constituent

	item	type
coverage	0.03	0.033
regularity	0.95	0.969

Support level: *supported*

Examples: Kosmetik**ø**industrie, Akustik**ø**paneele, Architektur**ø**büro, Metzgerei**ø**produkt
Counterexamples: Kultur**en**folge, Melodie**n**reigen, Partei**en**staat

Sources: Ortner and Müller-Bollhagen (1991, 107)

Constraint (31)

First constituents that are constituted by polysyllabic feminine nouns that end with [t] often attach *-s-* if the [t] is not part of a suffix *-(ig)keit*, *-heit*, *-schaft*, or *-ität* or its allomorphs.

P2L | MDP | first constituent

	item	type
coverage	0.013	0.016
regularity	0.505	0.463

Support level: *partially supported*

Examples: Arbeit**s**tag, Heirat**s**antrag, Zukunft**s**angst
Counterexamples: Arbeit**ø**geber, Jugend**ø**alter, Umwelt**ø**schutz

Sources: Nübling and Szczepaniak (2013, 77), Ortner and Müller-Bollhagen (1991, 74, 75)

Constraint (32)

First constituents that are constituted by nouns that end in schwa mostly adopt *-(e)n-*.

P2L | P | first constituent

	item	type
coverage	0.172	0.149
regularity	0.752	0.72

Support level: *supported*

Examples: Biene**n**zucht, Suppe**n**schüssel, Auge**n**lied
Counterexamples: Fläche**ø**maß, Sorge**ø**pflicht, Gebäude**ø**komplex, Gedanke**ns**turm

Sources: Kürschner (2010, 18), Ortner and Müller-Bollhagen (1991, 83, 92)

Constraint (33)

First constituents that are constituted by feminine nouns that end in schwa adopt *-(e)n-* regularly.

P2L | MP | first constituent

	item	type
coverage	0.156	0.133
regularity	0.765	0.744

Support level: *not supported*

Examples: Biene**n**zucht, Suppe**n**schüssel, Tiefe**n**meter
Counterexamples: Fläche**ø**maß, Sorge**ø**pflicht

Sources: Eisenberg (2013, 229), Fuhrhop and Kürschner (2015, 573), Krott et al. (2007, 3), Libben et al. (2009, 151, 157), Ortner and Müller-Bollhagen (1991, 83, 92), Schäfer and Pankratz (2018, 9)

Constraint (34)

First constituents that are constituted by consonant-final weak feminine nouns mostly adopt *-ø-* or *-s-* in compounds in which the second constituent forces a singular meaning of the given first constituent.

P2L | MPS | first constituent

Examples: Schriftøführer, Geburt**s**tag, Burgøanlage
Counterexamples: Burg**en**blick, Frau**en**bild

Sources: Eisenberg (2013, 229), Krott et al. (2007, 3), Nübling and Szczepaniak (2013, 80), Ortner and Müller-Bollhagen (1991, 94, 110)

Constraint (35)

First constituents that are constituted by consonant-final weak feminine nouns — especially final-stressed incl. monosyllabic ones — attach *-(e)n-* more likely in compounds in which the second constituent forces a plural meaning of the given first constituent.

P2L | MPS | first constituent

Examples: Schrift**en**verzeichnis, Geburt**en**kontrolle, Burg**en**land
Counterexamples: Burg**en**blick

Sources: Eisenberg (2013, 229), Nübling and Szczepaniak (2008, 3), Nübling and Szczepaniak (2013, 80), Ortner and Müller-Bollhagen (1991, 110), Schäfer (2018, 29)

Constraint (36)

Copulative compounds insert *-ø-* regularly. By copulative compounds are understood compounds in which the two constituents do not exhibit a clear modifier-head relation.

P2L | S | first + second constituent

Examples: Dichterøcomponist, Bettøsofa, Königinømutter

Sources: Koliopoulou (2014, 63, 64), Neef and Borgwaldt (2012, 31, 32), Ortner and Müller-Bollhagen (1991, 57, 91, 94, 109)

Constraint (37)

Argumental compounds are sometimes marked by *-s-*. By argumental compounds are understood compounds in which the second constituent still contains a high degree of verbiness and the first constituent of which constitutes their argument. The second constituent is such compounds is usually an agent designation, an *-ung* formation etc.

P2L | S | first + second constituent

Examples: Gewicht**s**heber, Krieg**s**führung, Unternehmen**s**beratung
Counterexamples: Gewichtøheber, Kriegøführung

Sources: Nübling and Szczepaniak (2013, 79)

Constraint (38)

Compounds belonging to technical terminology (economics, law, medicine, etc.) are often missing an explicit linking element.

P2L | L | first + second constituent

Examples: Erbschaftøsteuer, Herzøschlagen, Schadenøersatz
Counterexamples: Schaden**s**ersatz

Sources: Nübling and Szczepaniak (2008, 11), Ortner and Müller-Bollhagen (1991, 98), Schlücker (2023, 20), Wegener (2005, 177)

Constraint (39)

Compounds whose first constituent designates a person and whose second constituent is one of 'Mann', 'Frau', 'Leute', 'Tochter', 'Gattin', or 'Witwe' tend to insert *-s-*.

P2L | SL | first + second constituent

Examples: Reiter**s**mann, Lehrer**s**tochter, Bäcker**s**leute

Sources: Nübling and Szczepaniak (2008, 11), Ortner and Müller-Bollhagen (1991, 56)

Constraint (40)

The probability of *-s-* grows irregularly with the morphological complexity of the first constituent (any form of derivation).

P2L | D | first constituent

	item	type
coverage	1.0	1.0
regularity	0.735	0.766

Support level: *supported*

Examples: Gesundheit**s**amt, Gesellschaft**s**politik, Anspruch**s**haltung, Eintrag**s**frist
Counterexamples: Ort**s**amt, Krieg**s**ende, Anruføbeantworter, Minderheit**en**recht

Sources: Fuhrhop and Kürschner (2015, 572), Kopf (2017, 201), Kürschner, 106, 111, 112, Kürschner (2010, 19, 22)

Constraint (41)

The probability of *-s-* grows irregularly with the phonological complexity of the first constituent. By phonologically complex words are understood words with polysyllable non-trochaic form, words with unstressed prefixes, words with stressed or semi-stressed suffixes etc.

P2L P first constituent

	item	type
coverage	1.0	1.0
regularity	0.723	0.761

Support level: *supported*

Examples: Anruf**ø**beantworter, Beruf**s**erfahrung, Religion**s**unterricht
Counterexamples: Schlaf**ø**plan, Herbst**ø**anfang, Bestand**ø**teil

Sources: Fuhrhop and Kürschner (2015, 572), Nübling and Szczepaniak (2008, 10, 21, 22), Nübling and Szczepaniak (2013, 78)

Constraint (42)

The probability of *-s-* in compounds with simplex first constituents tends to decrease with the increasing sonority of the final segment of the first constituent. *-s-* is thus more frequent after plosives, infrequent after nasals and liquids, and it never occurs after a full vowel.

P2L P first constituent

	item	type
coverage	0.431	0.498
regularity	0.796	0.717

Support level: *supported*

Examples: Ort**s**tarif, Glück**s**rad, Himmel**ø**reich, Uhu**ø**paar
Counterexamples: Arbeit**ø**geber, Dieb**ø**stahl, Himmel**s**bahn, Uhu**s**nest

Sources: Fuhrhop and Kürschner (2015, 572), Kürschner, 110, Nübling and Szczepaniak (2008, 7, 17), Schäfer and Pankratz (2018, 9), Wegener (2003, 434), Wegener (2005, 180, 181, 182)

Constraint (43)

Almost all simplex nouns that constitute first constituents that attach *-s-* belong to a small fixed group of masculine and neuter nouns.

L2P M first constituent + linker

	item	type
coverage	0.05	0.032
regularity	0.988	0.915

Support level: *supported*

Examples: Zwang**s**jacke, Ort**s**angabe, König**s**haus
Counterexamples: Arbeit**s**amt, Heirat**s**antrag, Liebe**s**brief

Sources: Kürschner (2010, 23), Nübling and Szczepaniak (2013, 79), Schäfer and Pankratz (2018, 8)

Constraint (44)

Many simplex nouns that constitute first constituents that attach *-s-* have a high token frequency.

L2P L first constituent + linker

	item	type
coverage	0.05	0.032
regularity	0.649	0.914

Support level: *supported*

Examples: Volk**s**brauch, Amt**s**anwalt, Staat**s**amt
Counterexamples: Fasching**s**foto, Gral**s**sage, Konvent**s**sitzung

Sources: Kürschner (2010, 25)

Constraint (45)

All feminine nouns that constitute first constituents that attach *-s-* are morphologically complex.

L2P MD first constituent + linker

	item	type
coverage	0.163	0.159
regularity	0.996	0.983

Support level: *supported*

Examples: Gesundheit**s**amt, Gesellschaft**s**politik, Liebe**s**brief
Counterexamples: Arbeit**s**tag, Heirat**s**antrag

Sources: Kürschner (2010, 23, 24)

Constraint (46)

Almost all feminine nouns that constitute first constituents that attach *-s-* are polysyllabic.

L2P MP first constituent + linker

	item	type
coverage	0.163	0.159
regularity	0.999	1.0

Support level: *supported*

Examples: Gesundheit**s**amt, Gesellschaft**s**politik, Heirat**s**antrag, Liebe**s**brief

Sources: Ortner and Müller-Bollhagen (1991, 73)

Constraint (47)

All nouns that constitute first constituents that attach the *-n-* allomorph of the *-(e)n-* linker end in schwa.

L2P P first constituent + linker

	item	type
coverage	0.116	0.109
regularity	0.959	0.985

Support level: *partially supported*

Examples: Biene**n**zucht, Suppe**n**schüssel, Auge**n**lied
Counterexamples: Oper**n**haus, Elster**n**nest

Sources: Kürschner (2010, 18), Nübling and Szczepaniak (2013, 84), Wegener (2005, 179)

Constraint (48)

All nouns that constitute first constituents that attach *-(e)n-* belong to weak nouns.

L2P | M | first constituent + linker

	item	type
coverage	0.143	0.127
regularity	0.971	0.986

Support level: *partially supported*

Examples: Blume**n**topf, Katze**n**fell, Frau**en**hand
Counterexamples: Hahn**en**kamm, Mond**en**schein, Instrument**en**bau

Sources: Kopf (2017, 187), Kürschner, 118, 119, Kürschner (2010, 9), Nübling and Szczepaniak (2008, 3), Nübling and Szczepaniak (2013, 79, 84)

Constraint (49)

All masculine nouns that do not build the plural form with *-(e)n* that constitute first constituents that attach *-(e)n-* are monosyllabic designations of male persons, animals, astronomic objects, months.

L2P | MPS | first constituent + linker

Examples: Hahn**en**kamm, Mond**en**schein, Stern**en**himmel, Mai**en**nacht
Counterexamples: Mondøaufgang, Sternøbild, Maiøbaum

Sources: Nübling and Szczepaniak (2013, 79, 80), Ortner and Müller-Bollhagen (1991, 77, 78)

Constraint (50)

All neuter nouns that do not build the plural form with *-(e)n* that constitute first constituents that attach *-(e)n-* are polysyllabic foreign nouns that have a stressed final syllable that exhibit a plural meaning within the compound.

L2P | MPLS | first constituent + linker

Examples: Instrument**en**bau, Dokument**en**sammlung
Counterexamples: Zertifikatøsystem

Sources: Ortner and Müller-Bollhagen (1991, 78, 79)

Constraint (51)

All nouns that constitute first constituents that attach *-e-* build the plural form with *-e*.

L2P | M | first constituent + linker

	item	type
coverage	0.023	0.008
regularity	0.982	0.994

Support level: *supported*

Examples: Tag**e**buch, Punkt**e**stand, Hund**e**leine
Counterexamples: Herz**e**leid, Maus**e**zahn

Sources: Kürschner (2010, 9, 18), Nübling and Szczepaniak (2008, 5), Nübling and Szczepaniak (2013, 70, 81), Ortner and Müller-Bollhagen (1991, 102)

Constraint (52)

All nouns that constitute first constituents that attach *-e-* have a stressed last syllable.

L2P | P | first constituent + linker

	item	type
coverage	0.023	0.008
regularity	0.973	0.992

Support level: *partially supported*

Examples: Tag**e**buch, Punkt**e**stand, Hund**e**leine
Counterexamples: Anhalt**e**stelle

Sources: Kürschner (2010, 9, 18)

Constraint (53)

Most nouns that constitute first constituents that attach *-e-* are simplex.

L2P | D | first constituent + linker

	item	type
coverage	0.023	0.008
regularity	0.73	0.783

Support level: *supported*

Examples: Tag**e**buch, Punkt**e**stand, Hund**e**leine
Counterexamples: Anhalt**e**stelle

Sources: Kürschner (2010, 9, 18), Ortner and Müller-Bollhagen (1991, 102)

Constraint (54)

All nouns that constitute first constituents that attach *-e-* are native (none are loanwords).

L2P | L | first constituent + linker

Sources: Ortner and Müller-Bollhagen (1991, 102)

Constraint (55)

All nouns that constitute first constituents that attach *-(")er-* build the plural form with *-(")er*.

L2P M first constituent + linker

	item	type
coverage	0.01	0.015
regularity	0.979	0.999

Support level: *supported*

Examples: B**ü**ch**er**regal, Kr**ä**ut**er**frau, Kind**er**jacke

Sources: Eisenberg (2013, 229), Kürschner (2010, 9, 11, 18), Nübling and Szczepaniak (2008, 3), Nübling and Szczepaniak (2013, 80, 81), Ortner and Müller-Bollhagen (1991, 98)

Constraint (56)

All nouns that constitute first constituents that attach *-(")er-* have a stressed last syllable.

L2P P first constituent + linker

	item	type
coverage	0.01	0.015
regularity	0.938	0.96

Support level: *partially supported*

Examples: B**ü**ch**er**regal, Kr**ä**ut**er**frau, Kind**er**jacke
Counterexamples: Mitglied**er**liste, Vorbild**er**sammlung

Sources: Kürschner (2010, 9, 11)

Constraint (57)

Most nouns that constitute first constituents that attach *-(")er-* are simplex.

L2P D first constituent + linker

	item	type
coverage	0.01	0.015
regularity	0.875	0.935

Support level: *supported*

Examples: B**ü**ch**er**regal, Kr**ä**ut**er**frau, Kind**er**jacke
Counterexamples: Mitglied**er**liste, Vorbild**er**sammlung

Sources: Ortner and Müller-Bollhagen (1991, 98)

Constraint (58)

All nouns that constitute first constituents that attach *-(")er-* are native (none are loanwords).

L2P L first constituent + linker

Examples: B**ü**ch**er**regal, Kr**ä**ut**er**frau, Kind**er**jacke

Sources: Ortner and Müller-Bollhagen (1991, 98)

Constraint (59)

All nouns that constitute first constituents that attach *-"e-* build the plural form with *-e* and umlaut.

L2P M first constituent + linker

	item	type
coverage	0.011	0.003
regularity	1.0	1.0

Support level: *supported*

Examples: H**ä**nd**e**druck, **Ä**rzt**e**streik, G**ä**st**e**buch

Sources: Eisenberg (2013, 228), Nübling and Szczepaniak (2008, 5)

Constraint (60)

All nouns that constitute first constituents that attach *-es-* belong to a fixed row of masculine and neuter nouns and build the genitive form with *-(e)s-*. *-es-* is thus isolated.

L2P LM first constituent + linker

Examples: Bund**es**tag, Kind**es**alter, Tag**es**schau, Land**es**regierung

Sources: Eisenberg (2013, 229), Kürschner (2010, 11, 18), Nübling and Szczepaniak (2008, 3, 7), Nübling and Szczepaniak (2013, 81), Schäfer and Pankratz (2018, 8)

Constraint (61)

All nouns that constitute first constituents that attach *-es-* are monosyllabic.

L2P P first constituent + linker

	item	type
coverage	0.012	0.011
regularity	0.914	0.9

Support level: *partially supported*

Examples: Bund**es**tag, Kind**es**alter, Tag**es**schau, Land**es**regierung
Counterexamples: Gesetz**es**änderung, Vorjahr**es**winner, Bestand**es**aufnahme

Sources: Kürschner (2010, 11, 18)

Constraint (62)

All nouns that constitute first constituents that attach *-(e)ns-* belong to a fixed group of a few masculine nouns and a single neuter noun and have different genitive endings. *-(e)ns-* is thus isolated.

L2P L first constituent + linker

Examples: Schmerz**ens**geld, Name**ns**tag, Herz**ens**angst

Sources: Nübling and Szczepaniak (2013, 82), Ortner and Müller-Bollhagen (1991, 80, 97), Schäfer and Pankratz (2018, 9)

Constraint (63)

All nouns that constitute first constituents that attach *-ø-* with umlaut build the plural form with a zero ending and umlaut.

L2P | M | first constituent + linker

	item	type
coverage	0.003	0.0
regularity	1.0	1.0

Support level: *supported*

Examples: Br**ü**derøgemeinde, M**ü**tterøzentrum, V**ä**terøaufbruch

Sources: Nübling and Szczepaniak (2008, 5)

Constraint (64)

Linkers that are identical to the plural ending of the first constituent in a given compound tend to be associated with a plural meaning of this first constituent within the compound.

L2P | MS | first constituent + linker

Examples: Punkt**e**stand, B**ü**ch**er**regal, Staat**en**bund, Burg**en**land, M**ü**tterøzentrum
Counterexamples: Motor**en**geräusch, Burg**en**blick, Kind**er**jacke, Hund**e**leine, Pilzøsammler, Buchøhandel, Vogeløfutter

Sources: Krott et al. (2007, 3), Schäfer (2018, 230, 231), Schlücker (2023, 20), Wegener (2005, 173)

F. Suggestions to Corrections of Constraint Definitions

This appendix enlists our suggestions to corrections of constraint definitions that are partially/not supported by the empirical data. Dedicated Python checkers for the corrected versions of constraints proved that all the corrected versions are fully supported by the corpus data.

Constraint (3)

First constituents that are constituted by nouns that build the plural form with *-s* attach a zero linker regularly.

Correction: Almost all first constituents that are constituted by nouns that build the plural form with *-s* attach a zero linker.

	original	corrected
quantifier	regularly	almost all
item regularity	0.97	0.97

Constraint (19)

First constituents that are constituted by derived nouns with suffixes *-bold*, *-nis*, *-rich*, *-at*, *-al* that build the plural form with *-e* attach a zero linker regularly.

Correction: First constituents that are constituted by derived nouns with suffixes *-bold*, *-nis*, *-rich*, *-al* that build the plural form with *-e* attach a zero linker regularly.

	original	corrected
quantifier	regularly	regularly
item regularity	0.747	1.0

Constraint (25)

First constituents that are constituted by deverbal nouns that end in suffix *-en* attach *-s-* regularly.

Correction: First constituents that are constituted by deverbal nouns that end in suffix *-en* mostly attach *-s-*. Zero linker variants with a significantly lower frequency may occur.

	original	corrected
quantifier	regularly	mostly
item regularity	0.773	0.797

Constraint (29)

First constituents that are constituted by nouns that end in a full vowel always adopt a zero linker.

Correction: First constituents that are constituted by nouns that end in a full vowel adopt a zero linker regularly unless this noun is a loanword that ends with a stressed long [i] (*-ie*) or a stressed long [e] (*-ee*).

	original	corrected
quantifier	always	regularly
item regularity	0.968	0.993

Constraint (31)

First constituents that are constituted by polysyllabic feminine nouns that end with [t] often attach *-s-* if the [t] is not part of a suffix *-(ig)keit*, *-heit*, *-schaft*, or *-ität* or its allomorphs.

Correction: First constituents that are constituted by polysyllabic feminine nouns that end with [t] sometimes attach *-s-* if the [t] is not part of a suffix *-(ig)keit*, *-heit*, *-schaft*, or *-ität* or its allomorphs.

	original	corrected
quantifier	often	sometimes
item regularity	0.505	0.505

Constraint (33)

First constituents that are constituted by feminine nouns that end in schwa adopt -n- regularly.

Correction: <DISCARD AND MERGE WITH (32): even adding a simplex condition that eliminates deverbal and deadjectival nouns from consideration results in item regularity of 0.868; there is therefore no point in distinguishing between feminine and non-feminine nouns ending in schwa since both end up in the moderate regularity level.> First constituents that are constituted by nouns that end in schwa mostly adopt -n-.

	original	corrected
quantifier	regularly	mostly
item regularity	0.765	0.765

Constraint (47)

All nouns that constitute first constituents that attach the *-n-* allomorph of the *-(e)n-* linker end in schwa.

Correction: All nouns that constitute first constituents that attach the *-n-* allomorph of the *-(e)n-* linker end in schwa or in [ə] (unstressed *-er*).

	original	corrected
quantifier	all	all
item regularity	0.959	0.988

Constraint (48)

All nouns that constitute first constituents that attach *-(e)n-* belong are weak nouns.

Correction: Almost all nouns that constitute first constituents that attach *-(e)n-* belong are weak nouns.

	original	corrected
quantifier	all	almost all
item regularity	0.971	0.971

Constraint (52)

All nouns that constitute first constituents that attach *-e-* have a stressed last syllable.

Correction: All nouns that constitute first constituents that attach *-e-* have a stressed last syllable or are derived of those with a stressed prefix.

	original	corrected
quantifier	all	all
item regularity	0.973	1.0

Constraint (56)

All nouns that constitute first constituents that attach *-(")er* have a stressed last syllable.

Correction: All nouns that constitute first constituents that attach *-(")er* have a stressed last syllable or are derived of those with a stressed prefix.

	original	corrected
quantifier	all	all
item regularity	0.938	0.979

Constraint (61)

All nouns that constitute first constituents that attach *-es-* are monosyllabic.

Correction: All nouns that constitute first constituents that attach *-es-* are monosyllabic or are derived of those with a prefix.

	original	corrected
quantifier	all	all
item regularity	0.914	0.983

Additionally, we provide corrections for the constraints that were supported by the corpus data but appeared to be even more regular than specified in the relevant literature.

Constraint (16)

First constituents that are constituted by mixed masculine and neuter nouns almost always attach *-s-*, a zero linker, or *-(e)n-*.

Correction: First constituents that are constituted by mixed masculine and neuter nouns regularly attach *-s-*, a zero linker, or *-(e)n-*.

	original	corrected
quantifier	almost always	regularly
item regularity	0.997	0.997

Constraint (22)

First constituents that are constituted by deadjectival feminine nouns with a schwa suffix almost always attach -(e)n- or a zero linker. Each of the two linkers is preferred in about the same number of cases.

Correction: First constituents that are constituted by deadjectival feminine nouns with a schwa suffix always attach -(e)n- or a zero linker. Each of the two linkers is preferred in about the same number of cases.

	original	corrected
quantifier	almost always	always
item regularity	0.97	0.97

Constraint (28)

First constituents that are constituted by nouns that end in a sibilant or in a consonant cluster including [s] mostly adopt a zero linker.

Correction: First constituents that are constituted by nouns that end in a sibilant or in a consonant cluster including [s] adopt a zero linker almost regularly.

	original	corrected
quantifier	mostly	almost regularly
item regularity	0.97	0.97

Constraint (40)

The probability of *-s-* grows irregularly with the morphological complexity of the first constituent (any form of derivation).

Correction: The probability of *-s-* exhibits a tendency to grow with the morphological complexity of the first constituent (any form of derivation).

	original	corrected
quantifier	irregularly	tendency
item regularity	0.735	0.735

Constraint (41)

The probability of *-s-* grows irregularly with the phonological complexity of the first constituent. By phonologically complex words are understood words with polysyllable non-trochaic form, words with unstressed prefixes, words with stressed or semi-stressed suffixes etc.

Correction: The probability of *-s-* exhibits a tendency to grow with the phonological complexity of the first constituent. By phonologically complex words are understood words with polysyllable non-trochaic form, words with unstressed prefixes, words with stressed or semi-stressed suffixes etc.

	original	corrected
quantifier	irregularly	tendency
item regularity	0.723	0.723

Constraint (46)

Almost all feminine nouns that constitute first constituents that attach *-s-* are polysyllabic.

Correction: All feminine nouns that constitute first constituents that attach *-s-* are polysyllabic.

	original	corrected
quantifier	almost all	all
item regularity	0.999	0.999