

ZWJW

Zeitschrift für Wortbildung Journal of Word Formation

herausgegeben von / edited by

Petra M. Vogel • Ludwig M. Eichinger
Mechthild Habermann • Jörg Meibauer
Barbara Schlücker • Hans-Joachim Solms
Pavol Štekauer • Salvador Valera Hernández



1/2025

ZWJW

**Zeitschrift für Wortbildung
Journal of Word Formation**

2025 • Jahrgang / Volume 9 • Heft / Issue 1

HRSGG. / EDS.

Petra M. Vogel

Ludwig M. Eichinger

Mechthild Habermann

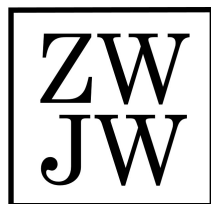
Jörg Meibauer

Barbara Schlücker

Hans-Joachim Solms

Pavol Štekauer

Salvador Valera Hernández



WISSENSCHAFTLICHER BEIRAT / ADVISORY BOARD:

Aleksandra Bagasheva (Sofia University)
Irmhild Barz (University of Leipzig)
Geert Booij (University of Leiden)
Jan Čermák (Charles University Prague)
Gerrit J. Dimmendaal (University of Cologne)
Jan Don (Open University Netherlands)
Nanna Fuhrhop (University of Potsdam)
Livio Gaeta (University of Turin)
Luise Kempf (University of Bern)
Lívia Körtvélyessy (Pavol Jozef Šafárik University Košice)
Elisabeth Leiss (Munich University)
Hans Christian Luschützky (University of Vienna)
Francesca Masini (University of Bologna)
Franz Rainer (Vienna University of Economics and Business)
Anna Thornton (University of L'Aquila)
Carola Trips (University of Mannheim)
Kristel Van Goethem (F.R.S.-FNRS & Université catholique de Louvain)
Martina Werner (Austrian Center for Digital Humanities)

REDAKTION / EDITORIAL TEAM:

Ambra Ottersbach & Carolin Gerwinski
Zeitschrift für Wortbildung / Journal of Word Formation
Universität Siegen
Fakultät I – Germanistik / Linguistik
Hölderlinstraße 3
D-57076 Siegen
E-Mail: zwjw@germanistik.uni-siegen.de
Homepage: journals.linguistik.de/zwjw

HOSTING PROVIDER:

Universitätsbibliothek Johann Christian Senckenberg
Bockenheimer Landstraße 134-138
D-60325 Frankfurt am Main
www.ub.uni-frankfurt.de

ISSN 2367-3877



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

Inhaltsverzeichnis / Table of Contents

BEITRÄGE / CONTRIBUTIONS

Martin Haspelmath

Compound and Incorporation Constructions as Combinations of
Unexpandable Roots.....1–23

Ivan Lacić

Competing Strategies in Morphological Approximation: Exploring Prefixoids
kvazi(-), nadri(-), nazovi(-), and pseudo(-) in Croatian.....24–71

Jan Seifert

Nutzende – Kritisierende – Forschende: Zum Gebrauch substantivierter
Partizipien als genderinklusive Referenzstrategie.....72–117

VARIA / MISCELLANEOUS

Melissa Mujkić

Tagungsbericht: „Perspektiven der Movierungsforschung“
(8.–10. Oktober 2024; Universität Luxemburg).....118–129

Martin Haspelmath

Compound and Incorporation Constructions as Combinations of Unexpandable Roots

Abstract: This paper offers definitions of the terms *compound construction*, *compound* and *incorporation construction* that can be applied to all languages in the same way. The earlier literature has often expressed pessimism about identifying such elements across languages in an objective way, but I propose definitions that do not rely on notions such as “word” or “morphology”. I define *incorporation* as a special kind of verbal compound construction, and a compound construction as a combination of strictly adjacent roots. Rather than being “non-phrasal”, I say that compounds are defined as not expandable by modifiers. These definitions are shared-core definitions, like most other definitions of comparative concepts: They capture the core of the types of elements that have been called “compound” and “incorporation” in the earlier literature, but not necessarily every compound and every incorporation in every language. It should be noted that the paper does not make any empirical claims, but merely contributes to the methodology of general linguistics by means of a critical and constructive discussion of terminology.

Keywords: compound, incorporation, root, comparative concept

1. Defining “Compound Construction” and “Incorporation Construction”

In this paper, I propose and discuss definitions of the well-known terms *compound* and *incorporation* in general grammar. Both these terms are well-known to linguists, have long been used and are associated with rich stereotypes, but so far, no commonly accepted definitions have emerged. I propose the definitions of *compound construction* and *incorporation construction* (as comparative concepts) that are given in (1) and (2). As can be seen, an incorporation construction is defined as a special kind of compound construction.

(1) **compound construction**

A compound construction is a construction consisting of two strictly adjacent slots for roots that cannot be expanded by full-nominal, adjectival, or degree modifiers.

(2) **incorporation construction**

An incorporation construction is an event-denoting verb-noun compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

Some examples of compounds from different languages are given in (3), and some examples of incorporating verb forms are given in (4).

(3) some compound forms¹

- | | | | |
|-------------------|---------------------|---------------|-------------------------------|
| a. German | <i>Auto-bahn</i> | [car-way] | ‘expressway’ |
| b. French | <i>tire-bouchon</i> | [pull-cork] | ‘corkscrew’ |
| c. Chinese | <i>fēi-jī</i> (飛機) | [fly-machine] | ‘airplane’ |
| d. Mwotlap | <i>tit teñteñ</i> | [punch cry] | ‘make (s.o.) cry by punching’ |
| e. Guinean Kpelle | <i>kwéli-kóyô</i> | [panther-leg] | ‘panther leg’ |

(4) some incorporating verb forms

- | | |
|----------------------------------|---|
| a. Mohawk
(Iroquoian) | <i>wak-tsi’tsia-ientho-on</i>
1SG-flower-plant-ST
‘I planted flowers’ (Mithun & Corbett 1999: 49) |
| b. Biniñ Gunwok
(Gunwinyguan) | <i>ba-warde-jobge-ng</i>
3PL-rock-split-PST
‘they split the rock’ (Evans 1996: 65) |
| c. Guarani
(Tupi-Guarani) | <i>ai-po-pete</i>
1SG-hand-slap
‘I slapped the hand’ (Velázquez-Castillo 1996: 99) |
| d. Mapudungun
(Araucanian) | <i>kintu-waka-le-y</i>
seek-cow-PROG-IND.SG
‘he is looking for the cows’ (Baker et al. 2004: 139) |
| e. Soninke
(Mande) | <i>Múusá hèri-kétí</i>
Moussa donkey-beat
‘Moussa beat donkeys’ (Creissels & Dramé 2018: (24b)) |

¹ The first three examples are from my own knowledge. Mwotlap is an Austronesian language of Vanuatu (the example is from François 2004: 112), and Kpelle is a Mande language of Guinea (the example is from Green & Konoshenko 2022: 12).

The definitions proposed here have the following notable properties that will be discussed in the course of this paper:

- (i) They can be applied equally to all languages as they do not make reference to language-particular features (§2).
- (ii) They do not presuppose a distinction between morphology and syntax (see Haspelmath 2011), and they are not based on the notion of ‘word’, so that ‘word’ can be defined with reference to ‘compound’ (Haspelmath 2023a).
- (iii) They are not prototype-based or fuzzy.
- (iv) They single out the great majority of constructions that have been called ‘compound’ and ‘incorporation’, as well as the most typical cases, but not all cases (i.e. they are shared-core definitions; §10).

It should be noted that the paper does not make any empirical claims, but merely contributes to the methodology of general linguistics by means of a critical and constructive discussion of terminology. Moreover, there is no claim that these concepts are particularly significant for linguistics, and especially the definition of *incorporation* is clearly so specific that there seems to be nothing natural about it. However, it is necessary to explain these terms to newcomers (e.g. in textbooks and encyclopedia articles) because they are so widely used, and it is best to provide clear definitions in order to avoid the impression that the terms necessarily correspond to natural categories of the world. Linguists often assume (sometimes implicitly) that grammar is by nature divided into morphology and syntax, but this view was inherited from a long tradition and is not a result of linguists’ research. Thus, it is best not to reinforce the stereotypical view of compounds as belonging to “morphology” (as opposed to “phrases” that belong to “syntax”), and to define both compounds and phrases as types of constructions (see Haspelmath 2023b for the term *construction*, which is neutral between “morphology” and “syntax”).

Thus, this paper is a proposal to improve the methodology of general linguistics by offering clear definitions for two technical terms that are widely used by linguists. Previous overview articles often note that clear definitions are unavailable or difficult (e.g. Schlücker 2023: §2.2: “There is no clear or general definition of compound”), but this state of affairs is not unavoidable. The definitions proposed here will not satisfy everyone, but as no

empirical claims are involved, this is not important. The central message is that we can have clear and relatively simple definitions of core technical terms of grammar, and that we do not need to wait until we have solutions to deep problems such as the purported syntax-morphology distinction.

2. Avoiding Language-Particular Defining Features

The definitions in (1)-(2) are intended as general concepts that can be applied to any language, and thus they do not contain concepts that are relevant only to particular languages. In a given language, a category with the label “compound” may well be defined and delimited differently. For example, German Compounds are defined with respect to a special stress pattern (e.g. Compound *Rótwein* ‘red wine’, contrasting with the Phrase *ròter Wéin* ‘wine which is red’),² but a stress criterion plays no role in most other languages. Modern Greek Compounds are defined with respect to the absence of inflectional affixes on the modifier root (Ralli 2013: 23), but again, this criterion cannot be applied to all languages as not all languages have such inflectional affixes.

The language-particular nature of the definitions has sometimes been associated with a difficulty of identifying compounds cross-linguistically:

- (5) a. Dressler (2006: 24)
“More explicit universal definitions of the intensional type are not only theory-dependent ... but also cross-linguistically never watertight – in many languages there are exceptions or fuzzy transitions to non-compounding.”
- b. Lieber & Stekauer (2009: 7)
“[The] criteria might have limited utility within a particular language or group of languages, but cross-linguistically they cannot be definitive.”
- c. Bauer (2017: 19)
“Unfortunately, the criteria are difficult to apply across language types.”
- d. Finkbeiner & Schlücker (2019: 9)
“In [defining compounds] we do not aim for more than a rough approximation, as it is clear that the respective criteria are not only in part language-specific, but also a matter of controversial theoretical debate.”

² Note that language-particular terms are written in upper case here, following a well-known convention.

Aikhenvald (2007: 24) puts it quite bluntly: “Compounds have to be defined on language-internal criteria”.

In this paper, by contrast, I propose defining criteria that are applicable across language types in the same way (without “difficulty”),³ and without fuzziness. The criteria are of course somewhat arbitrary, but this is almost always the case when choosing comparative concepts for general linguistics.

3. Compounds vs. Compound Constructions

The definitions in (1)–(2) characterize kinds of compound constructions, but we often also talk about compounds, i.e. kinds of forms. Thus, we should add a definition of *compound*:

(6) **compound**

A compound is a form (consisting of two adjacent roots) that instantiates, or was created by, a compound construction.

For regularly formed compounds, the distinction between “instantiation” and earlier “creation” is not relevant. For example, English regular compounds like *snake poison* or *apple cake*, or ad hoc compounds like *cow tree* (‘tree that cows like to rub up against’; Downing 1977: 827) were created from the construction $[N_2 - N_1]$ ‘ N_1 which is related to N_2 ’, and they can be said to currently instantiate this construction because their meanings are compatible with the general semantic schema.

However, there are many compounds with non-compositional meanings which cannot be said to instantiate a compound construction, e.g. *soap opera* (not a kind of opera), *honeymoon* (not a kind of moon). However, it seems clear that they were created from the same construction as regular compounds, and the figurative nature of their meanings is often still apparent. They are thus included by the condition “was created by” of the definition, as long as one can be sure that they were created by the same construction. Excluding non-compositional forms from the definition of compounds would give unintuitive results, but this is not a very important point (as unintuitive results cannot be avoided in general; this is discussed in §10 below).

³ Actually, I would say that defining technical terms is never difficult (see my blogpost about this: Haspelmath 2022). What is often difficult is to find a clear definition of an old term that applies to all languages in the same way and that will make one’s colleagues happy.

One could object that the definition in (6) includes a diachronic criterion (“or was created from”) that should not have a place in definitions. This is a valid point, and it would alternatively be possible to define a compound simply as a form that instantiates a compound construction. The definition would then be much narrower, but it would still correspond largely to our intuitions.

4. Compounds Consist of Roots (Not Stems or Words)

It is sometimes said that compounds consist of two words (Marchand 1960: 11) or two stems (Schlücker 2023), but some authors prefer to remain vague: Thus, Olsen (2015: 364) says that compounding involves the combination of “two or more lexemes (roots, stems, or freely occurring words)”. By contrast, my definition relies on the notion of root, in line with definitions such as (7).

- (7) “A simple way to make new lexemes is to make compounds by combining noun, verb and adjective roots.” (Gebhardt 2023: 133)

A root is defined as a non-composite form that has content meaning (rather than function meaning), or more precisely as in (8) (for further discussion, see Haspelmath 2025).⁴

- (8) **root**
A root is a contentful morph (i.e. a morph denoting an action, an object or a property) that can occur as part of a free form without another contentful morph.

Restricting compounds to root combinations means that the verbal prefixes of Indo-European languages (e.g. German *um-armen* [around-arm] ‘embrace’, Russian *vy-nimat’* [out-take] ‘take out’) do not count as compound elements, which is in line with current usage.⁵ It also means that combinations involving pronouns (e.g. English *him-self*) or adpositions (e.g. *on-to*) are not regarded as compounds, which again corresponds to the way the term *compound* is generally used.

It is not felicitous to define compounds as consisting of two words or two lexemes because words often include inflectional affixes, and we generally think of compound

⁴ Note that *root* is not defined in terms of ‘word’, but it is the other way round: For the definition of *word*, we need the more basic concept ‘root’ (Haspelmath 2023a).

⁵ As observed by Olsen (2015: §1.1), the older traditional practice from the 19th and earlier 20th centuries (up to Marchand 1960) was to regard such forms as compounds rather than as prefixed formations.

stems as not including inflectional affixes. Compounds may of course be inflected by adding further affixes, but the two compound constituent elements are roots. In languages with thoroughgoing case and number inflection, the compound members do not carry such affixes. For example, the Latin adjective-noun compound *magn-anim-(us)* ‘great-spirited’ includes the root *magn-* ‘great’, but not any of the case and number inflection that normally occurs on the adjective (nominative singular *magn-us*, accusative singular *magn-um*, dative plural *magn-is*, and so on).⁶

Or should we say that compounds consist of *stems* rather than roots? This would have the advantage that commonly found forms such as those in (9) would be included, where one of the elements is a derived form rather than a root (*free-dom*, *Heiz-ung*, *es-és*).

- | | | | | | |
|-----|----|-----------|---------------------|------------------|---------------|
| (9) | a. | English | <i>free-dom day</i> | | |
| | b. | German | <i>Gas-heiz-ung</i> | [gas-heat-ing] | ‘gas heating’ |
| | c. | Hungarian | <i>hó-es-és</i> | [snow-fall-NMLZ] | ‘snowfall’ |

But what exactly is a “stem”? There is a good definition of the term *root* (as given in (8) above), but we do not have a generally accepted definition of *stem*.⁷ One could suggest that a stem is a combination of a root and derivational affixes, but this presupposes a definition of “derivational” as opposed to “inflectional”, and so far, there is no generally accepted way of drawing the line between these (but see Haspelmath 2024a for a recent proposal). As the notion of “stem” adds further complications, I restrict the definition to roots here.

Finally, it should be noted that the restriction to roots means that phrases cannot be compound members. Thus, so-called “phrasal compounds” (e.g. Trips & Kornfilt 2015) of the type *chicken and egg situation* do not fall under the current definition. They are of course compound-like in languages in which they occur, but there does not seem to be a way of defining *compound* in general such a way that they can be included.

⁶ “Co-compounds” such as Russian *čaški-bljudca* [cups-plates] ‘dishes’ (see Wälchli 2015) and “appositive compounds” such as Russian *inženier-fizik* [engineer-physicist] ‘person who is both engineer and physicist’ (Ohnheiser 2019: 256) would thus not count as compounds by the current definition, as they show case and number inflection of both members.

⁷ There is a proposal in Haspelmath (2025: n. 6), but the definition is not widely accepted and more complex than ‘root’.

5. The Roots Are Adjacent

If two roots are combined but are not always adjacent to each other, we do not call the resulting pattern “compounding”. Thus, the German expression *statt-finden* ‘take place’ in (10a) does not qualify as a compound, because when the verb is finite as in (10b), the noun-like part *statt* (‘place’) is not adjacent to the verb *findet*. (Thus, Wurzel’s (1996) suggestion that this pattern might be an instance of incorporation is not in line with the current definition.)

- (10) a. *Die Buchmesse kann dieses Jahr statt-finden.*
 the book.fair can this year place-take
 ‘The book fair can take place this year.’
- b. *Die Buchmesse finde-t dieses Jahr statt.*
 the book.fair take-s this year place
 ‘The book fair takes place this year.’

Since the definition is not meant to be a fuzzy concept or prototype (see (iii) in §1 above), this also means that constructions with linking elements as in (11) are not included.

- (11) not compounds
- a. German *Liebe-s-brief* [love-LK-letter] ‘love letter’
- b. English *bird’s nest*
- c. French *chemin de fer* [way of iron] ‘railway’
- d. M. Greek *vrox-ó-nero* [rain-LK-water] ‘rainwater’⁸

In Indo-European languages, such compound-like constructions are not uncommon, and at least in German, they have always been included in the (language-particular) category of Noun-Noun Compounds. In a classic article, Benveniste (1966) said that forms of the type *chemin de fer* are the true compounds of French. However, the subsequent literature has not adopted this way of talking about such Romance expressions. They have recently been called “phrasal lexemes” (Masini 2009) or *binominal lexemes* (Masini et al. 2023; Pepper 2023), defined in terms of the classifying or naming function of such forms. However, the term *compound* is generally defined in a strictly formal way (see, e.g., the definitions listed by Scalise & Vogel 2010: 5), and this tradition is followed here. As a result,

⁸ This example is from Ralli (2013: 47); the relevant words are *vroxí* ‘rain’ and *neró* ‘water’.

not only Romance expressions containing a preposition of the *de/di* type are excluded, but also Germanic expressions that include a genitive-type marker or other additional markers that are not part of either of the roots.

Another example of a compound-like construction that is not a compound is shown in (12). One might consider treating *lí-hūn* [leave-marriage] ‘divorce’ as an incorporation (see Wang 2022), but some adverbials can intervene between the two parts of this composite expression. It is thus more similar to English expressions like *take part*, which cannot be considered incorporations either because some affixes can intervene (*take-s part*, *tak-ing part*).

(12) Mandarin Chinese

- | | | | |
|----|-----------------------------------|-------------------------------|------------------|
| a. | <i>lí-hūn</i> (離婚) | [leave-marriage] | ‘divorce’ |
| b. | <i>lí-le-liǎng-ci-hūn</i> (离了两次婚) | [leave-PRF-two-time-marriage] | ‘divorced twice’ |

6. Roots in Compounds Cannot Be Expanded

We now get to the key property of compound constructions that distinguishes them from “syntactic” constructions: nonexpandability. A widely shared intuition is that compounds are formed “morphologically”, and they are typically included under the heading of “word formation”. But how can this be translated into a straightforward criterion that can be applied uniformly to all languages? This question is often formulated as the question of “distinguishing between compounds and phrases” (e.g. Schlücker & Plag 2011; Ralli 2013: 243–268; Cetnarowska 2019: 15–44; Gebhardt 2023: 136–140).

But how do we recognize a “phrase”? In sentences such as *cats like milk*, or *he lacks courage*, the subject and object forms are typically treated as nominal phrases rather than simply words, but this cannot be seen in these particular example sentences. The reason why we say that *cats*, *milk*, *he* and *courage* occupy phrasal slots in a construction of the type [NP V NP] is that they can be expanded by articles and adjectival or nominal modifiers, as in *small cats like my neighbour’s milk*, or *he lacks the necessary courage*. For adjective-noun compounds, the impossibility of modification by degree adverbs has often been noted in the literature; for example, Finkbeiner & Schlücker (2019: 10) observe that the first element in German *Alt-bau* [old-building] cannot be expanded by a degree adverb (**sehr Alt-bau* ‘very old building’).

For Modern Greek, Ralli (2013: 21) notes that the phrase *áyria yáta* ‘wild cat’ can be expanded by a noun phrase modifier or by coordination, while the constituents of the compound *ayrió-yata* ‘wildcat’ cannot be expanded in this way (though it is of course possible to modify the entire compound, as in *meyáli ayrió-yata* ‘big wildcat’).⁹ This is illustrated in (13)–(14).

- (13) a. *áyria yáta*
wild cat
‘wild cat’
- b. *i áyria tis Mariás i yáta*
the wild of Maria the cat
‘Maria’s wild cat’
- c. *áyria ke meyáli yáta*
wild and big cat
‘wild and big cat’
- (14) a. *ayrió-yata*
wild-cat
‘wildcat (*Felis silvestris*)’
- b. **ayrio-mavrió-yata*
wild-black-cat
‘wild black cat’
- c. **poli-ayrió-yata*
very-wild-cat
‘very wild cat’

In Guinean Kpelle, a Mande language, the two construction types differ only in their tonal properties (Green & Konoshenko 2022: 12). When a possessor nominal (e.g. a noun expanded by a postnominal numeral) precedes, the modified noun does not undergo a tonal change (e.g. *kóyó* ‘leg’ in (15)). In the compound construction in (16a) (= 3e), the modified noun undergoes a tonal change (*kóyô*). This tonal form is not possible when the modifying element is expanded as in (16b).

⁹ It should be noted that the expansion criterion does not distinguish fixed phrases from fixed compounds, but only freely formed phrases from freely formed compounds. For example, Schlücker & Plag (2011: 1541) observe that the German fixed phrase *schwarzes Brett* [black board] ‘bulletin board’ cannot be expanded to **sehr schwarzes Brett* [very black board], but this is for semantic reasons. (Thus, Gebhardt’s **very bluebird* is not a fully pertinent example, because *bluebird* is a fixed idiomatic compound.)

- (15) *kwéli* *hààbá* *kóyó*
panther three leg
'legs of three panthers'

- (16) a. *kwéli-kóyô* (= 3e)
panther-leg
'panther leg'
- b. **kwéli hààbá kóyô*

The definition in (1) does not mention articles or coordination because the definition can make use only of universally applicable concepts, and articles and coordination do not exist in all languages. As adjectival modifiers and possessive modifiers of nouns occur in all languages, and likewise degree modifiers of adjectives (presumably) occur in all languages, they are suitable for the general definition. For verbs, many languages have adjectival modifiers (e.g. English *walk slowly*, German *schön singen* ‘sing beautifully’), and maybe one should add other types of expansions to the definition in (1) to distinguish verb roots in compounds from verbs occurring in “phrases”.

Thus, the criterion of expandability allows us to distinguish between compounds and what has traditionally been called “phrases” without requiring a definition of “phrase”. We will see below in §9 that this criterion also allows us to distinguish incorporation from constructions that are not incorporation.

7. Compounds Need Not Have a Naming or Generic Function

The most typical nominal compounds denote established concepts that are known to the language users independently of the compound construction. For example, *summer vacation* in (17a) is an established concept in modern Western culture, while (17b) refers to a specific instance of a summer vacation experienced by a specific person. Schlücker & Plag (2011: 1539) say that compounds are “inherently suitable for kind reference (or “naming”), due to their status as word formation entities”, and this seems to be a widespread view.

- (17) a. *summer vacation* (naming compound)
b. *last summer's vacation in Czechia* (specific phrase)

However, compounds need not have a classifying or naming function, or refer to kinds rather than specific referents. Many languages allow ad hoc compounds such as *office vacation* in (18a), which is not an established concept but could refer to an unusual kind of vacation (e.g. a vacation during which an agricultural worker gets to use an empty office and can relax by working on their memoir for a few weeks). And in addition, compounds need not be generic (kind-referring), but the modifying root can refer to a specific person, as in *Mitterrand interview* in (18b).

- (18) a. *office vacation* (ad hoc compound)
 b. *the Mitterrand interview* (specific modifying root)

As I noted above, *compound* and *incorporation* are best defined in strictly formal terms, and frequently occurring functions of compound or incorporation constructions are best characterized by different terms. This is also the conclusion of Schlücker & Plag (2011), who note that not only adjective-noun compounds (such as *Rotwein* ‘red wine’) can have a naming function in German, but also adjective-noun phrases (such as *großer Zeh* ‘big toe’ or *bunter Abend* [colorful evening] ‘evening of music and entertainment’).¹⁰ Croft (2022: 141–142) calls the naming or generic function *typefying*, and he notes that many languages do not use compounds for typefying constructions (e.g. preferring binominals such as French *chemin de fer* ‘railway’, cf. §5 above). A compound construction is thus a construction-strategy, not a construction-function.

8. Some Further Properties of the Definition of Compound

Before moving on to incorporation constructions, I will briefly comment on five further topics: (i) the specification of two roots; (ii) the treatment of “synthetic compounds”; (iii) the treatment of “neoclassical compounds”; (iv) the possible presence of an external compound marker; and (v) the order of the two roots.

¹⁰ The precise nature of the “naming function” of compounds has not been widely discussed by semanticists, as far as I am aware. A plausible suggestion may be that this function falls under the notion of *clichés*, which is discussed by Mel’čuk (2015), who treats it as a subtype of phraseme. Clichés are composite expressions which are compositional but whose use is not fully predictable from their parts and their constructional meanings. They are thus inventorial items even though they are not irregular (see Haspelmath 2024b on *inventorization* as a type of “lexicalization”).

First, one may ask why the definition restricts compound constructions to binary combinations of roots. There is no strong reason for this, but it makes the definition more concrete than saying “several roots”, and the possible ternary (or n -ary for $n > 2$) combinations that are excluded in this way seem to be marginal. One might say that a compound such as *Brazil-Argentina game* is ternary (because *Brazil-Argentina* is not an element that undisputably exists outside of the compound), but such compounds are uncommon, and semantically they are not ternary.

Second, what about “synthetic compounds” such as German *hart-herz-ig* [hard-heart-y] ‘uncharitable, hard-hearted’ (based on *hart* and *Herz*), or Spanish *medi-ev-al* [middle-age-al] ‘medieval’ (based on *medi(-o)* and *ev(-o)*)? These include two roots, but crucially an additional affix, so they are not compounds according to the present definition.

Third, “neoclassical compounds” of European languages such as *geo-logy*, *bio-graphy*, *demo-cracy* do not count as compounds because their components are not roots (in that they do not occur in free forms without another root-like form; see (7)).

Fourth, as noted by Trips & Kornfilt (2015), Turkish compound-like forms include a marker at the end of the second root, e.g. *müzik festival-i* ‘music festival’. Such markers go beyond what is specified in the definition, so they do not count as compounds. One might ask what they are if they are not compounds, because they are not ordinary phrases either. But this is not a question that I address here. Not every language-particular construction must fall under some well-known general concept, and there are many *sui generis* phenomena in languages.

And fifth, the definition does not specify the order in which the two roots occur, so theoretically, the order could be free. I do not know about compound constructions with free order, but if there are no such constructions, this may be regarded as a testable universal, and it need not be specified in the definition.

There are no deep reasons for all of these decisions, and some of them have been taken in order to make the definition simpler. As I noted earlier, such definitions must be somewhat arbitrary, because it is impossible to formulate a definition that applies to all languages equally and at the same time singles out the language-particular classes.

9. More on Incorporation

As I noted in §1, this paper defines incorporation as a subtype of compounding. The definition in (2) above is repeated here for convenience.

(2) **incorporation construction**

An incorporation construction is an event-denoting verb-noun compound construction in which the noun occupies an argument slot of the verb and occurs in a position where nominal patient arguments cannot occur.

Analogously to the distinction between *compound construction* and *compound* (in §3 above) we might distinguish between an *incorporation construction* and an *incorporation* (as a kind of form that instantiates an incorporation construction), though this is not usual in the literature.¹¹

That the verb-noun compounds that are called incorporations refer to the event denoted by the verb is the first important condition, needed to exclude compounds with verb modifiers such as English *push-cart* or German *Wasch-maschine* ‘washing machine’, as well as exocentric compounds such as French *tire-bouchon* [pull-cork] ‘corkscrew’. In addition, the noun must occupy an argument position of the verb (mostly that of patient argument).

Let us briefly look at a few definitions of (*noun*) *incorporation* in the literature, to see some of the specificities of the present definition:¹²

- (19) a. “Noun incorporation is the compounding of a noun stem and a verb (or adjective) to yield a complex form that serves as the predicate of a clause.” (Gerdt 1998)
- b. [noun incorporation:] “a construction in which a noun stem is combined with a verb to form a new, morphologically complex verb” (Mithun 2000: 916)

¹¹ Note that *incorporation* is more often used to denote an abstract process, analogously to *compounding* (= the abstract process of forming a compound). In the present paper, I avoid such abstract process nouns.

¹² Note that I do not talk about “noun incorporation” but simply about *incorporation*, which is by definition restricted to noun-verb combinations of a particular type. Some authors have used expressions such as “adverb incorporation”, and occasionally one reads about incorporation of other elements, but these phenomena are clearly marginal and can safely be ignored here. Here I follow Gerdt (1998), who basically defines “incorporation” as coextensive with compound formation and thus treats noun incorporation as a special case of it (see also Haugen 2015: 414). (For completeness, I note that “incorporated pronouns” are a very different matter, not closely related to traditional incorporation: When person indexes are affixed, they are sometimes said to be “incorporated”, in a usage that goes back to Sapir (1911: 250) or earlier, but this terminological usage is confusing.)

- c. “We will define noun incorporation as instances where a bare or reduced nominal displays a close linear linguistic relation with a verb, through either morphology or strict adjacency, and where the head of this unit is clearly verbal in its distribution or marking.” (Johns 2017: §1)
- d. “As a starting point, we can say that it refers to a grammatical phenomenon whereby a nominal element, usually with an internal thematic role, forms some kind of unit with a verbal element, and together they serve as the verb or predicate of a sentence.” (Massam 2017: §1)
- e. “Incorporation can be described as the inclusion of one lexical element in another lexical element such that they together constitute a single word.” (Olthof 2020: 131)
- f. “Noun incorporation prototypically consists of a verbal compound consisting of a verbal root and a nominal root.” (Barrie & Mathieu 2020: 265)

Most of these definitions are largely compatible with the present definition, but some of them make use of concepts that are not needed here: “word” in (19e), “morphology” in (19b, c), “predicate” in (19a, d), “head” in (19c), “lexical element” in (19e). I do not know how to define these terms, so the present definition of *incorporation* instead makes reference to well-defined notions such as *compound* (instead of “morphology”), *root* (instead of “lexical element”), and “event-denoting” (instead of “head”).

Incorporation became well-known in the 1980s because of a prominent controversy between Mithun (1984; 1986) representing a “lexical” view of incorporation and Sadock (1986) representing a “syntactic” view, and because of Baker’s (1988) use of the term “incorporation” for an abstract syntactic operation (see Haugen 2015 for an account of key aspects of those debates). Like many hotly debated issues in linguistics, it has not been resolved, and if one does not presuppose a distinction between “syntax” and “lexicon”, the controversy loses much of its interest.

One issue has been the question whether an incorporated noun can be referential and serve as an antecedent for a subsequent anaphoric construction. Sadock (1986) notes that this is possible in Greenlandic, as seen in (20a–b).

(20) Greenlandic (Sadock 1986: 23)

- a. *Luutiviup assut qusanartumik qaanniorpaa,*
 Luutivik-p assut qusanartoq-mik qajaq-lior-paa
 Luutivik-ERG very beautiful-INS kayak-make.for-INDIC.3SG>3SG
 ‘Luutivik made him a very beautiful kayak,’

- b. *unnermillu amertillugu*.
 unneq-mik=lu amertit-lugu
 white.sealskin-INS=CONJ cover-CONTEMP.3SG
 ‘and covered it with white sealskin.’

Sadock argued that this shows that at least some languages have a kind of noun incorporation that is syntactic, and that not all incorporations are “lexical” (as claimed by Mithun 1984). However, this observation does not affect the current discussion, because (i) there is no claim that parts of compounds cannot be specific or referential (see §7),¹³ and (ii) the Greenlandic construction discussed by Sadock does not count as incorporation by the criteria of the present definition. This is because Greenlandic *-lior/-nior* ‘make for’ (as in *qaan-nior-paa* ‘made a kayak for him’) is a suffix rather than a root. As specified in the definition in (8), a root “can occur as part of a free form without another contentful morph”, but this is not the case for *-lior/-nior*, which always occurs following a noun root and is thus a derivational suffix, not a root. Eskimo languages are unusual in having a large number of derivational suffixes (often called “postbases”), and it is also unusual that the noun roots in such derived verbs can be referential, but these are not cases of incorporation.

Likewise, derivational suffixes that mean ‘make’ or ‘have’ in Uto-Aztecan languages are not roots, although they have been treated in the context of incorporation by Hale & Keyser (2002) and Haugen (2008) (see also Haugen 2015: §4). Two examples from O’odham are given in (21).

- (21) Tohono O’odham (Uto-Aztecan; Hale & Keyser 2002: 133)
- | | | | | |
|----|-------------|----------|---------------|-----------------|
| a. | <i>hoa</i> | ‘basket’ | <i>hoa-t</i> | ‘make a basket’ |
| b. | <i>si:l</i> | ‘saddle’ | <i>si:l-t</i> | ‘make a saddle’ |

Such denominal suffixes are also discussed by Gerdts & Marlett (2008) and contrasted with incorporation. These authors note that the Halkomelem element *tx^w*- ‘buy’ cannot occur on its own, but only in combination with a noun root as in (22a). Thus, this construction is not an incorporation construction, and the element *tx^w*- is a verbalizing prefix.

¹³ See also Barrie & Mathieu (2016: 4) for a similar argument from referentiality for a “syntactic” view of incorporation (on the basis of data from Algonquian and Iroquoian). They simply work with a different understanding of the term “incorporation” (and they do not even assume that “incorporation” is a unified phenomenon).

(22) Halkomelem (Salishan; Gerdts & Marlett 2008: 411)

- a. *Ni? cən tɬʷ-səplil.*
 AUX 1.SBJ VBL-bread
 ‘I bought bread.’
- b. **Ni? cən tɬʷ-ət kʷ səplil.*
 AUX 1.SBJ buy-TR DET bread
 ‘I bought some bread.’
- c. *Ni? cən ʔiləq-ət kʷ səplil.*
 AUX 1.SBJ buy-TR DET bread
 ‘I bought some bread.’

Finally, the definition in (2) specifies that the incorporated noun must occur in a position different from the position of the argument when it is a full nominal. For example, in Turkish, bare caseless nouns in object function can only occur directly in front of the verb, as in (23a), in contrast to accusative-marked objects, as seen in (23b).

(23) Turkish (Aydemir 2004: 465–466)

- a. *Yasemin anahtar kaybet-ti.* (**Anahtar Yasemin kaybet-ti.*)
 Yasemin key lose-PST
 ‘Yasemin lost keys (or a key).’
- b. *Anahtar-ı Yasemin kaybet-ti.*
 key-ACC Yasemin lose-PST
 ‘Yasemin lost the key.’ (OR: ‘The key was lost by Yasemin.’)

So does (23a) represent a case of noun incorporation, as hinted by Aydemir (2004)? I would say that it does not, because the corresponding definite nominal can occur in the same position (*Yasemin anahtarı kaybet-ti*, which is basically equivalent to (23b)), so it does not occur in a special position where a full-nominal patient object cannot occur.

10. Shared-Core Definitions of Concepts for General Linguistics

For some or many readers, one of the most surprising or perhaps irritating aspects of the current proposal is that several types of constructions or expressions that are typically treated as compounds are not covered by the definitions. In particular, the exclusion of compound-like forms with composite components and of “phrasal compounds” (§4), as well as the exclusion of “synthetic compounds” (§8), may seem unintuitive. However, it must be kept in mind that such “surprising” properties are a necessary feature of all

definitions of general (comparative) grammatical terms, because they are defined in a way different from language-particular categories. General concepts cover the SHARED CORE of the extensions of the comparable categories of different languages.

For example, the English Present Tense and the German Present Tense are not coextensive, because only the German Present Tense can regularly be used for future time reference as well (e.g. *wir kommen morgen* ‘we will come tomorrow’), but they both match the comparative concept “present tense”. Or the Turkish Dative case and the Russian Dative case are not coextensive, because only the Turkish Dative case can be used for spatial goals as well (e.g. *Moskova’ya* ‘to Moscow’), but they both match the comparative concept “dative case”. Or the French Feminine gender class is not coextensive with the English Feminine gender class because only the French Feminine gender class includes many inanimate nouns (e.g. *la lune* ‘the moon’), but they both match the comparative concept “feminine gender class”.

Similarly, German Compounds and Chinese Compounds overlap in a core set of phenomena that match the definition in (1), but for both languages, there are language-particular criteria that lead researchers to include more phenomena. This is not different from the language-particular criterion for French Feminine gender, which is defined by the article *la*, not by the meanings of the nouns. The extent to which the shared-core definition covers the domains of the various language-particular categories is a matter of convention, and here I have proposed one particular choice. Comparative concepts are always somewhat arbitrary, and for traditional terms, it seems best to choose reasonably simple definitions.¹⁴

One reviewer asks: If compound-like forms such as “phrasal compounds” and German Compounds with linking elements are not included in the present definition of the term *compound*, what are they instead? This may seem like a logical question, but in the present paper, I discuss general terms for general-comparative linguistics. Such concepts are intended to compare languages, not to describe them. Language description must be exhaustive, but comparative approaches do not extend to the totality of a language. One can easily make up new general concepts and terms that might extend to German

¹⁴ One reviewer says that “the more comprehensive the definition is the better”, but this is not generally the case. Definitions of traditional terms should not be too complex, because otherwise they cannot be used in textbooks. This may mean that they are less comprehensive than one might expect.

Liebe-s-brief ‘love letter’, but here the purpose is merely to offer a definition of the traditional term *compound* that may be used for textbook purposes.

11. Conclusion

To conclude, the definitions of *compound* and *incorporation* proposed in this paper make it possible to compare these phenomena across languages without presupposing a distinction between syntax and morphology, or definitions of “word” or “phrase”. A key concept here is that of expansion by full-nominal, adjectival or degree modifiers, which yields results that earlier linguists often tried to describe in terms of a “word–phrase” distinction or a “syntax–morphology” distinction.

In what sense does this constitute progress? There is no empirical contribution in this paper, but the definitions suggested here show that there is no reason to resign oneself to a situation where we do not have a definition shared by the discipline. Quite a few authors have expressed rather sobering or negative views about definitions, e.g.

- (24) a. “There is no overall agreement on such basic issues as the definition of a compound. Accordingly, there can be no agreement on whether compounding is a linguistic universal or not.” (Bauer 2017: 1–2)
- b. “It’s difficult to classify compounds and phrases into two distinct morphosyntactic structures.” (Gebhardt 2023: 140)
- c. “Noun incorporation can be defined in a number of ways.” (Johns 2017: §1)
- d. “There is a lot of disagreement about exactly what constitutes noun incorporation.” (Massam 2017: §1)

It seems that the main reason why linguists often disagree about the precise meanings of even rather basic general terms is that they think of these terms as denoting aspects of nature, rather than concepts created by linguists in order to compare languages. If they were natural kinds, one would not need a proper definition (perhaps only a “working definition”), because one would arrive at the correct notion once enough research has been done. Linguists often seem to assume that a good definition of a term is something that will

result from a definitive theory of the domain in question, or at least that our definitions will get better as our theories get better.¹⁵

However, if general terms of linguistics denote comparative concepts rather than natural kinds, there is no reason to wait for better theories. There does not seem to be any tendency for linguists' general-theoretical and methodological views and preferences to converge over the decades, so it is worth exploring the possibility of finding definitions that do not rely on controversial theoretical views and that could be understood widely (e.g. for textbook purposes). It is hoped that the present definitions have this property, so that even if readers are not happy to adopt them, they will appreciate the fact that the definitions do not rely on other basic concepts (such as “syntax vs. morphology”, or “head”, or “phrase”) that are themselves controversial or poorly understood. The present definition of ‘compound construction’ also provides a basis for a definition of ‘word’ in the same spirit (Haspelmath 2023a). It remains to be seen whether these definitions help us make progress in theoretical understanding, but they should at least help us avoid the kinds of confusions that we often saw in the past.

References

- Aikhenvald, Alexandra. 2007. Typological distinctions in word-formation. In Timothy Shopen (ed.), *Language typology and syntactic description, Vol. 3*, 1–65. Cambridge: Cambridge University Press.
- Aydemir, Yasemin. 2004. Are Turkish preverbal bare nouns syntactic arguments? *Linguistic Inquiry* 35(3). 465–474.
- Baker, Mark C., Roberto Aranovich & Lucía A. Golluscio. 2005. Two types of syntactic noun incorporation: Noun incorporation in Mapudungun and its typological implications. *Language* 81(1). 138–176.
- Baker, Mark C. 1988. *Incorporation: A theory of grammatical function changing*. Chicago: University of Chicago Press.
- Barrie, Michael & Eric Mathieu. 2016. Noun incorporation and phrasal movement. *Natural Language & Linguistic Theory* 34(1). 1–51. DOI: [10.1007/s11049-015-9296-6](https://doi.org/10.1007/s11049-015-9296-6).
- Barrie, Michael & Eric Mathieu. 2020. Noun incorporation and polysynthesis. In Daniel Siddiqi, Michael Barrie, Carrie Gillon, Jason D. Haugen & Eric Mathieu (eds.), *The Routledge handbook of North American languages*, 253–274. New York, NY: Routledge.

¹⁵ A characteristic statement is the following by Downing (1977: 810): “I have declined to adopt any of these criteria, since I doubt that the dividing line between N + N compounds and nominal phrases is always well-defined”. One can infer from this statement that Downing does not regard it as her task as a linguist to provide a definition, and that instead she thinks that whether such a definition can be found is a research question.

- Bauer, Laurie. 2017. *Compounds and compounding*. Cambridge: Cambridge University Press.
- Benveniste, Émile. 1966. Formes nouvelles de la composition nominale. *Bulletin de la Société de Linguistique de Paris* 61(1). 82–95.
- Cetnarowska, Bożena. 2019. *Compound nouns and phrasal nouns in English and Polish*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Creissels, Denis & Djibril Dramé. 2018. Transitivity and incorporation in Soninke. *Frankfurter Afrikanistische Blätter* 26. 37–54
- Creissels, Denis & Djibril Dramé. 2018. Transitivity and incorporation in Soninke. In Rose-Juliet Anyanwu (ed.), *Transitivity in African languages*, 37–54. Cologne: Köppe. http://www.denis-creissels.fr/public/Creissels_Drame-IncorpSon.pdf
- Croft, William. 2022. *Morphosyntax: Constructions of the world's languages*. Cambridge: Cambridge University Press.
- Downing, Pamela. 1977. On the creation and use of English compound nouns. *Language* 53(4). 810–842. DOI: [10.2307/412913](https://doi.org/10.2307/412913).
- Dressler, Wolfgang U. 2006. Compound types. In Gary Libben & Gonia Jarema (eds.), *The representation and processing of compound words*, 23–44. Oxford: Oxford University Press.
- Evans, Nicholas. 1996. The syntax and semantics of body part incorporation in Mayali. In Hilary Chappell & William McGregor (eds.), *The grammar of inalienability*, 65–110. Berlin: Mouton de Gruyter.
- Finkbeiner, Rita & Barbara Schlücker. 2019. Compounds and multi-word expressions in the languages of Europe. In Barbara Schlücker (ed.), *Complex lexical units: Compounds and multi-word expressions*, 1–44. Berlin: De Gruyter. DOI: [10.1515/9783110632446-001](https://doi.org/10.1515/9783110632446-001).
- François, Alexandre. 2004. Chains of freedom: Constraints and creativity in the macro-verb strategies of Mwotlap (Vanuatu). In Isabelle Bril & Françoise Ozanne-Rivierre (eds.), *Complex predicates in Oceanic languages*, 107–143. Berlin: Mouton de Gruyter.
- Gebhardt, Lewis. 2023. *The study of words: An introduction*. London: Routledge. DOI: [10.4324/9781003030188](https://doi.org/10.4324/9781003030188).
- Gerds, Donna B. & Stephen A. Marlett. 2008. Introduction: The form and function of denominal verb constructions. *International Journal of American Linguistics* 74(4). 409–422. DOI: [10.1086/595571](https://doi.org/10.1086/595571).
- Gerds, Donna B. 1998. Incorporation. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 84–100. Oxford: Blackwell.
- Green, Christopher R. & Maria Konoshenko. 2022. Tonal head marking in Mande compounds: endpoint neutralization and outliers. *Mandenkan: Bulletin semestriel d'études linguistiques mandé* (67). 3–44. DOI: [10.4000/mandenkan.2699](https://doi.org/10.4000/mandenkan.2699).
- Hale, Ken & Samuel Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, MA: MIT Press.
- Haspelmath, Martin. 2011. The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1). 31–80. DOI: [10.1515/flin-2017-1005](https://doi.org/10.1515/flin-2017-1005).
- Haspelmath, Martin. 2022. Defining is never “difficult” – the practical problem is the polysemy of terms. *Diversity Linguistics Comment* (blogpost). DOI: [10.58079/nswm](https://doi.org/10.58079/nswm).
- Haspelmath, Martin. 2023a. Defining the word. *WORD* 69(3). 283–297. DOI: [10.1080/00437956.2023.2237272](https://doi.org/10.1080/00437956.2023.2237272).
- Haspelmath, Martin. 2023b. On what a construction is. *Constructions* 15(1). DOI: [10.24338/cons-539](https://doi.org/10.24338/cons-539).

- Haspelmath, Martin. 2024a. Inflection and derivation as traditional comparative concepts. *Linguistics* 62(1). 43–77. DOI: [10.1515/ling-2022-0086](https://doi.org/10.1515/ling-2022-0086).
- Haspelmath, Martin. 2024b. Four kinds of lexical items: Words, lexemes, inventorial items, and mental items. *Lexique* (to appear). <https://ling.auf.net/lingbuzz/007880>.
- Haspelmath, Martin. 2025. Roots and root classes in comparative grammar. In Nataliya Levkovych, Julia Nintemann & Maike Vorholt (eds.), *Exploring structures in languages and language contact*. Berlin: De Gruyter Mouton (to appear).
- Haugen, Jason D. 2008. *Morphology at the interfaces: Reduplication and noun incorporation in Uto-Aztecan*. Amsterdam: Benjamins.
- Haugen, Jason D. 2015. Incorporation. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-Formation: An International Handbook of the Languages of Europe, Volume 1*, 413–434. Berlin: De Gruyter Mouton. DOI: [10.1515/9783110246278](https://doi.org/10.1515/9783110246278).
- Johns, Alana. 2017. Noun incorporation. In Martin Everaert & Henk van Riemsdijk (eds.), *The Wiley Blackwell companion to syntax, 2nd Edition*, 1–27. Wiley Blackwell. DOI: [h10.1002/9781118358733.wbsyncom093](https://doi.org/10.1002/9781118358733.wbsyncom093).
- Lieber, Rochelle & Pavol Štekauer. 2009. Introduction: Status and definition of compounding. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of compounding*, 3–18. Oxford: Oxford University Press.
- Lovestrand, Joseph. 2021. Serial verb constructions. *Annual Review of Linguistics* 7(1). 109–130. DOI: [10.1146/annurev-linguistics-031920-115317](https://doi.org/10.1146/annurev-linguistics-031920-115317).
- Marchand, Hans. 1960. *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Wiesbaden: Harrassowitz.
- Masini, Francesca, Simone Mattioli & Steve Pepper. 2023. Exploring complex lexemes cross-linguistically. In Steve Pepper, Francesca Masini & Simone Mattioli (eds.), *Binominal lexemes in cross-linguistic perspective: Towards a typology of complex lexemes*, 1–19. Berlin: De Gruyter Mouton.
- Masini, Francesca. 2009. Phrasal lexemes, compounds and phrases: A constructionist perspective. *Word Structure* 2(2). 254–271. DOI: [10.3366/E1750124509000440](https://doi.org/10.3366/E1750124509000440).
- Massam, Diane. 2017. Incorporation and pseudo-incorporation in syntax. *Oxford Research Encyclopedia of Linguistics* 2017. DOI: [10.1093/acrefore/9780199384655.013.190](https://doi.org/10.1093/acrefore/9780199384655.013.190).
- Mel'čuk, Igor. 2015. Clichés, an understudied subclass of phrasemes. *Yearbook of Phraseology* 6(1). 55–86. DOI: [10.1515/phras-2015-0005](https://doi.org/10.1515/phras-2015-0005).
- Mithun, Marianne & Greville G. Corbett. 1999. The effect of noun incorporation on argument structure. In Lunella Mereu (ed.), *Boundaries of morphology and syntax*, 49–71. Amsterdam: Benjamins.
- Mithun, Marianne. 1984. The evolution of noun incorporation. *Language* 60(4). 847–894.
- Mithun, Marianne. 1986. On the nature of noun incorporation. *Language* 62(1). 32–37.
- Mithun, Marianne. 2000. Incorporation. In Geert Booij, Christian Lehmann & Joachim Mugdan (eds.), *Morphology: A handbook on inflection and word Formation, Vol. 1*, 916–929. Berlin: Mouton de Gruyter.
- Ohnheiser, Ingeborg. 2019. Compounds and multi-word expressions in Russian. In Barbara Schlücker (ed.), *Complex lexical units: Compounds and multi-word expressions*, 251–278. Berlin: De Gruyter.

- Olsen, Susan. 2015. Composition. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe*, Vol. 1. Berlin: De Gruyter Mouton. DOI: [10.1515/9783110246254-017](https://doi.org/10.1515/9783110246254-017).
- Olthof, Marieke. 2020. Formal variation in incorporation: A typological study and a unified approach. *Linguistics* 58(1). 131–205. DOI: [10.1515/ling-2019-0036](https://doi.org/10.1515/ling-2019-0036).
- Pepper, Steve. 2023. Defining and typologizing binominal lexemes. In Steve Pepper, Francesca Masini & Simone Mattiola (eds.), *Binominal lexemes in cross-linguistic perspective: Towards a typology of complex lexemes*, 23–72. Berlin: De Gruyter Mouton.
- Ralli, Angela. 2013. *Compounding in Modern Greek*. New York: Springer.
- Sadock, Jerrold M. 1986. Some notes on noun incorporation. *Language* 62(1). 19–31.
- Sapir, Edward. 1911. The problem of noun incorporation in American languages. *American Anthropologist* 13(2). 250–282.
- Scalise, Sergio & Irene Vogel (eds.). 2010. *Cross-disciplinary issues in compounding*. Amsterdam: Benjamins.
- Schlücker, Barbara & Ingo Plag. 2011. Compound or phrase? Analogy in naming. *Lingua* 121(9). 1539–1551. DOI: [10.1016/j.lingua.2011.04.005](https://doi.org/10.1016/j.lingua.2011.04.005).
- Schlücker, Barbara. 2023. Compounding and linking elements in Germanic. *Oxford Research Encyclopedia of Linguistics* 2023. DOI: [10.1093/acrefore/9780199384655.013.954](https://doi.org/10.1093/acrefore/9780199384655.013.954).
- Spencer, Andrew. 1995. Incorporation in Chukchi. *Language* 71(3). 439–489.
- Trips, Carola & Jaklin Kornfilt. 2015. Typological aspects of phrasal compounds in English, German, Turkish and Turkic. *STUF: Language Typology and Universals* 68(3). 281–321. DOI: [10.1515/stuf-2015-0015](https://doi.org/10.1515/stuf-2015-0015).
- Velázquez-Castillo, Maura. 1996. *The grammar of possession: Inalienability, incorporation and possessor ascension in Guarani*. Amsterdam: Benjamins.
- Wälchli, Bernhard. 2015. Co-compounds. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe*, Vol. 1, 707–727. Berlin: De Gruyter Mouton. DOI: [10.1515/9783110246278](https://doi.org/10.1515/9783110246278).
- Wang, Yong. 2022. From syntax to morphology: Noun-incorporation in Chinese. *Studies in Language* 46(4). 872–900. DOI: [10.1075/sl.21015.wan](https://doi.org/10.1075/sl.21015.wan).
- Wurzel, Wolfgang Ullrich. 1996. Morphologischer Strukturwandel: Typologische Entwicklungen im Deutschen. In Lang, Ewald & Zifonun, Gisela (eds.), *Deutsch - typologisch*, 492–524. Berlin: Mouton de Gruyter. DOI: [10.1515/9783110622522-022](https://doi.org/10.1515/9783110622522-022).

Martin Haspelmath

Max Planck Institute for Evolutionary Anthropology

Department of Linguistic and Cultural Evolution

Deutscher Platz 6

04103 Leipzig

martin_haspelmath@eva.mpg.de



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

Ivan Lacić

Competing Strategies in Morphological Approximation: Exploring Prefixoids *kvazi(-)*, *nadri(-)*, *nazovi(-)*, and *pseudo(-)* in Croatian*

Abstract: This study explores the phenomenon of affix rivalry within the domain of morphological approximation in Croatian, focusing on the prefixoids *kvazi(-)*, *nadri(-)*, *nazovi(-)*, and *pseudo(-)* as they attach to nominal bases. These prefixoids can be classified as privative, as the derivatives they produce do not fully embody the core characteristics conveyed by their morphological bases. To analyze the rivalry among the prefixoids, the study evaluates their productivity, collocational behavior, and distribution across various textual genres, utilizing data from the *CLASSLA-web.hr* corpus. The findings suggest significant disparities in the productivity and collocational behavior of the prefixoids, with *nazovi(-)* and *kvazi(-)* exhibiting the highest productivity and highly overlapping collocational behavior, whereas *pseudo(-)* and *nadri(-)* reveal more specialized usage patterns. Additionally, a random sample of 500 tokens per prefixoid is annotated for semantic values. Again, *nazovi(-)* and *kvazi(-)* demonstrate substantial overlap, particularly in their mutual application as means for subjective depreciative evaluation, underscoring the insufficiency or pretentiousness of the subject. *Nadri(-)* is more narrowly focused on legal domains, while *pseudo(-)*, with its proclivity for scientific contexts, remains distinct but conceptually adjacent to *kvazi(-)* in contexts where imitation is highlighted without necessarily invoking deceit. Overall, the prefixoids present a complex network of interrelationships, yet each prefixoid also establishes a specific niche, balancing between shared semantic roles and distinct, context-dependent uses.

Keywords: affix rivalry, approximation, Croatian, evaluative morphology, prefixoids

1. Introduction

1.1. Approximation in Morphology

Morphological approximation represents a relatively underexplored area within the broader category of evaluative morphology. Evaluative morphology encompasses a diverse range of constructions that serve various semantic functions, all of which pertain to the

* I wish to acknowledge Jurica Polančec for his assistance with the preliminary data extraction. I would also like to thank two anonymous reviewers for their valuable comments. Any remaining errors are, of course, my own.

concept of linguistic subjectivity. Subjectivity is central to evaluation, seen as “a mental operation assessing the value of an object or event as more or less desirable and important from the interpreter’s perspective” (Merlini Barbaresi 2015: 38). Here, the speaker makes a judgment about value rather than stating factual information. In evaluative contexts, the standard of comparison is a mental construct, an abstract representation shaped by cultural and/or social factors and commonly shared within a community. As such, it is expected that different speakers may evaluate the same object, action, or person in various ways (Grandi & Körvélyessy 2015; Grandi 2017). While functions such as diminution, augmentation, and intensification have been extensively studied (cf., *inter alia*, Dressler & Merlini Barbaresi 1994; Grandi 2002; Körvélyessy & Štekauer 2011; Napoli & Ravetto 2017; for Croatian, although primarily on non-morphological means of intensification, see Batinić, Kresić & Pavić Pintarić 2015; Nigoević 2020; Lacić 2024a,b), morphological approximation has only recently begun to receive attention (Amiot & Stosic 2022; Masini, Norde & Van Goethem 2023). The relative paucity of research on approximation within morphological studies is highlighted by the variety of terms and lack of consensus on their definitions. For instance, Grandi & Körvélyessy (2015) describe this functional domain using the triad *approximation/reduction/attenuation*, while other terms include *deintensification* (Körvélyessy 2015), *non-prototypicality* (Cúneo 2015), as well as *non-authenticity*, *fakeness*, or *imitation* (Masini & Micheli 2020). These varying terms underscore the core concept of approximation as a comparison or resemblance to a concept based on specific properties. Crucially, approximative formations convey that category X, despite its similarities, is ultimately not fully present as it is lacking one or more of its prototypical features. As outlined by Masini, Norde & Van Goethem (2023), sources of approximation values are numerous, spanning over negation items, degree and quantity items, diminutives, simulative items, taxonomic items, etc. For this study, of a particular interest are markers conveying meanings of fakeness/imitation/pretending, most frequently expressed by *pseudo(-)*, derived from Ancient Greek *pseudēs* ‘false’. Such markers are characterized by their “privative” reading, implying that “(a) fake X is not (an) X” (Capelle, Denis & Keller 2018: 9). For instance, a *pseudodoctor* lacks the qualifications of a genuine doctor, even though it may share certain traits with them. Markers of fakeness/imitation/pretending, both cross-linguistically and intra-linguistically, are numerous (e.g., Van Goethem & Norde (2020)

explore eight Dutch “fake” morphemes), and the situation is not different for Croatian. In addition to the aforementioned *pseudo(-)*, Croatian uses two near-synonymous prefixoids of Slavic origin, namely *nadri(-)* (< IMP.2SG of verb *nadrijeti* ‘to begin’¹) and *nazovi(-)* (< IMP.2SG of verb *nazvati* ‘to name, to call’). Additionally, it is possible to join to that group also the prefixoid *kvazi(-)* (lat. *quasi* ‘almost’), which primarily conveys a meaning of incompleteness, in line with the prototypical approximation reading. However, we argue (see §3.4) that in Croatian it can also convey the concept of fakeness, just like the aforementioned three prefixoids. With four prefixoids occupying overlapping semantic niches, it becomes apparent that defining specific values of these prefixoids can be challenging due to their capacity to convey multiple, often similar meanings, resulting in affix rivalry. At its most basic, rivalry refers to “the fact that speakers routinely have to make a choice between alternative ways of realizing a certain concept” (Gardani, Rainer & Luschützky 2019: 4). This redundancy prompts linguistic systems to resolve competition through either specialization, akin to Aronoff’s (2019) *habitat niche differentiation*, or the elimination of a form altogether (Bauer, Valera & Díaz-Negrillo 2010). However, as Nagano, Bagasheva & Renner (2024: 3) point out, in word-formation the competition often simply continues since “(i) coexistence rather than disappearance is commonly observed, and (ii) the form of specialization tends to deviate from the elsewhere distribution [cf. Aronoff 2023].”

To the best of our knowledge, no contemporary studies have specifically addressed either the broader concept of morphological approximation in Croatian nor the competitive dynamics among rival forms. This study seeks to address this gap by examining four rival Croatian prefixoids prone to privative use, viz. *kvazi(-)*, *nadri(-)*, *nazovi(-)*, and *pseudo(-)*, as a member of <PREFIX_{APPRX} + noun> construction as exemplified by (examples taken from CLASSLA-web.hr):

- (1) *Mi vodimo svoje male i bezvredne živote, [...] a cijela armija umišljenih hrvatskih kvaziintelektualaca dobro i predobro živi na naš račun.*

‘We lead our small and worthless lives, [...] while a whole army of pretentious Croatian pseudo-intellectuals (lit. KVAZIintellectuals) lives well and too well at our expense.’

¹ It is noteworthy that in contemporary Croatian the verb *nadrijeti* is mostly used with a meaning of *navaliti*, *nahrupiti*, ‘rush in, come/enter suddenly’ (Šonje 2000).

- (2) *No, utjecaj ljevičarskih **nadri-intelektualaca** na svakodnevnu politiku sve je jači i na Zapadu [...].*

‘However, the influence of the left-wing pseudo-intellectuals (lit. NADRIintellectuals) on everyday politics is growing stronger in the West as well [...].’

- (3) *Upravo zato počeli smo pisati Studentsku deklaraciju [...] kako bi svijet čuo i glas hrvatskih studenata, a ne samo probranih, **nazovi intelektualaca**.*

‘This is precisely why we started writing the Student Declaration [...], so that the world could hear the voice of Croatian students, not just that of the selected, pseudo-intellectuals (lit. NAZOVI intellectuals).’

- (4) *Dok su ukus moderne umjetnosti sve više usmjeravali **pseudo intelektualci**, dotle se naivna umjetnost [...] razvijala sama.*

‘While the taste of modern art was increasingly directed by pseudo-intellectuals, naive art [...] developed independently.’

Naturally, approximation in Croatian can also be conveyed by other morphological processes (e.g., suffixation with *-ić* as in *doktorčić* ‘doctor.DIM’, which, among others, can have a reading synonymous to one with *nadri(-)* or *nazovi(-)*) or by syntactic means (e.g., adjective *tobožnji* ‘so-called/supposed’, as in *tobožnji liječnik*, ‘a supposed doctor’). This study, however, focuses on the competition among the four prefixoids discussed, which can be seamlessly integrated into the same syntactic structures without necessitating modifications in sentence constructions.

1.2. Previous Accounts

As previously noted, reference works concerning morphological approximation in Croatian are non-existent and beside basic dictionary descriptions of the four prefixoids, little is known. As far as dictionaries are concerned, two most comprehensive dictionaries were examined, viz. the *Dictionary of the Croatian language* (Šonje 2000) and VRH – *Large Dictionary of Croatian Standard Language* (Jojić 2015). Furthermore, a database *Croatian LanguagePortal*² (HJP), encompassing data from several Croatian dictionaries, was consulted. In Šonje (2000), there are no mentions of *kvazi(-)*, while for *pseudo(-)* six entries are listed, but only *pseudoklasicizam* ‘pseudoclassicism’ is relevant for this analysis, as the others pertain to scientific names of animal species. The prefixoid *nadri(-)* is represented by

² Available at: <https://hjp.znanje.hr/> (accessed 19 January 2025).

four lemmas: *nadriliječništvo* ‘quackery’, *nadriliječnik* ‘quack’, *nadriobrtnik* ‘quack craftsman’, and *nadripisar* ‘quack scribe’. All three profession denoting nouns describe individuals who untruthfully perform an activity for which they do not have the necessary education and competence. Along with the three profession denoting nouns, synonymic versions with *nazovi(-)* are reported. Finally, for *nazovi(-)*, Šonje (2000) reports only *nazoviliječništvo* ‘quackery’ and *nazovipisarstvo* ‘scribal quackery’, with no *nadri(-)* formations listed as synonyms. In contrast, VRH (Jojić 2015) includes entries for all four prefixoids. For *kvazi(-)*, we find prefixoids *pseudo(-)*, *nadri(-)*, and *nazovi(-)* mentioned as synonyms, but also adverbs *gotovo* ‘almost’ and *skoro* ‘nearly’. The following formations are found: *kvazijunak* ‘pseudo-hero’, *kvaziliteratura* ‘pseudo-literature’, and *kvaziliteraran* ‘pseudo-literary’. For *pseudo(-)*, no synonymic affixes are listed but we do find two nouns, viz. *pseudocivilizacija* ‘pseudo-civilization’ and *pseudočinjenica* ‘pseudo-fact’, as well as the adjective *pseudocivilizacijski* ‘regarding a pseudo-civilization’. All three definitions stress the fakeness of the modified head. Moreover, for *nadri(-)*, we find *nazovi(-)* as a synonymous formant, along with profession denoting nouns *nadriliječnik* ‘quack’, *nadripisar* ‘quack scribe’, *nadriobrtnik* ‘quack craftsman’, *nadripjesnik* ‘quack poet’ and *nadriumjetnik* ‘quack artist’, with the respective feminine pairs and derived relational adjectives, such as *nadriliječnički* ‘regarding a quack doctor’ and *nadripisarski* ‘regarding a quack scribe’. Furthermore, for *nazovi(-)*, VRH reports nouns *nazovijunak* ‘pseudo-hero’, *nazoviliječnik* ‘pseudo-doctor’, *nazoviliteratura* ‘pseudo-literature’, *nazoviobrtnik* ‘pseudo-craftsman’, *nazovipisar* ‘pseudo-scribe’, and, lastly, *nazovipjesnik* ‘pseudo-poet’. The final resource examined, the *Croatian Language Portal* (HJP), includes entries for all four prefixoids, but does not provide additional insight into the distinctions between them. Similar to VRH, *kvazi(-)* is explained with the prefixoids *pseudo(-)*, *nadri(-)*, and *nazovi(-)*, as well as adverbs *gotovo* ‘almost’ and *skoro* ‘nearly’, *tobože* ‘allegedly’ and *navodno* ‘allegedly’. Regarding *pseudo(-)*, the concept of fakeness and falsity is emphasized, and the example *pseudodemokracija* ‘pseudo-democracy’ is provided. The prefixoids, *nadri(-)*, *nazovi(-)*, *lažno(-)* ‘fake’, and *krivo(-)* ‘lit. wrong’ are listed as synonyms. For *nadri(-)*, HJP reports *nazovi(-)* as a synonymous formant, along with examples such as *nadriliječnik* ‘quack’ and *nadriknjiga* ‘quack book’. *Nadri(-)* is defined as conveying concepts of allegation and fakeness, equal to those activated by *nazovi(-)*, for which *nadri(-)* is listed as a synonymous

prefixoid. An example of *nazovi(-)* is given with the noun *nazoviprijatelj* ‘so-called friend’. In conclusion, the existing dictionary descriptions of these prefixoids, where available, are often circular and unhelpful for disambiguation. Each prefixoid is frequently defined using another synonymous prefixoid, reinforcing the shared semantics among them rather than clarifying their individual meanings.

Outside of dictionaries, the primary discussions surrounding these prefixoids are limited to debates on the categorical status of the formations they produce. The earliest available study on this topic, concerning *nadri(-)* is a work by Stjepan Ivšić (1906–1907) entitled *Nešto o riječima složenima s nadri-* [Something about words composed with *nadri-*]. Ivšić (1906–1907) observes that formations with *nadri(-)* were created analogously to the term *nadriknjiga*, denoting a person that has only superficially engaged with a book, thus referring to someone with a minimal grasp of a subject. By analogy, states Ivšić, the term *nadriliječnik* ‘quack’ was expected to indicate a doctor who is just starting their education/practice, but it does not mean that. Instead, the element *nadri(-)* is used in the sense of fake, self-proclaimed, with *nadriliječnik* ‘quack’ meaning ‘a person without the necessary schooling and professional qualification who performs the duties of a doctor’ (Šonje 2000). For that motive, Ivšić suggests that formations with *nadri(-)* should be replaced by *nazovi(-)* (e.g., *nazoviliječnik* ‘quack’). However, as later pointed out by Barić (1979, 1980), Ivšić has missed to notice that the meaning of the formations with *nadri(-)* and *nazovi(-)* is equivalent and *nazovi(-)* does not provide a more accurate alternative, as it does not preserve the original meaning. Barić (1980) concludes that *nadri(-)* and *nazovi(-)* act as modifiers of the word they attach to, imparting a meaning of fakeness, pretenses, and classifies them as prefixes. Finally, Barić also mentions *laži(-)* (IMP.2SG of verb *lagati* ‘to lie’), but noted its rarity compared to the others³. Returning to *nadri(-)*, Klajn (2002) concurs with Ivšić (1906–1907) that today

³ A research based on several contemporary Croatian dictionaries revealed only formations *lažitorba* (lit. LAŽIbag) ‘one who lies a lot’, and *lažidoktor* (lit. LAŽIdoctor) ‘a fake doctor, a person pretending to be a doctor’ (Jojić 2015). However, a brief search in the CLASSLA.web-hr corpus revealed over ten nominal formations with *laži-*, as well as two adjectival ones. Since to the best of our knowledge there is no previous work devoted to *laži-*, we deemed useful to report the found formations in a hope they will motivate some future research. Found nominal formations, besides the aforementioned *lažitorba* and *lažidoktor*, are: *lažibogomolja* ‘fake place of worship’, *lažibumbar* ‘fake bumblebee’, *lažijezik* ‘fake language’ (here it is curious to notice that *laži* in the formation *politički lažijezik* ‘political fake language’ keeps its original meaning and, in that given context, the formation refers to the communication full of lies, not to a fake,

nadri(-) is completely opaque and has lost every connection with the verb *nadrijeti* ‘to begin’ from which it is derived. Like Barić (1980), Klajn emphasizes that *nadri(-)* functions as a prefix rather than a component of a compound, paralleling the usage of borrowed prefixes such as *pseudo(-)* and *kvazi(-)*. On the other hand, discussing the categorical status of elements with *nazovi(-)*, Klajn (2002) states that, since the connection with the verb *nazvati* ‘to name, to call’ is completely transparent, words like *nazoviprijatelj* ‘pseudo-friend’ should be classified as compounds.

In summary, the review of existing research on Croatian prefixoids reveals a predominant focus on the two Slavic prefixoids, with earlier studies primarily concerned with their morphological status. These studies have largely neglected the semantic properties and competitive dynamics involving these prefixoids, both amongst themselves and in relation to non-native forms like *kvazi(-)* and *pseudo(-)*. In contrast, English-language studies provide some comparative insights into *quasi(-)* and *pseudo(-)*. Bauer, Lieber & Plag (2013) argue that while both elements share the characteristic of forming derivatives that do not represent genuine exemplars of their categories, their distinction lies in the element of falseness, which is inherent in *pseudo(-)* but absent in *quasi(-)*. Similarly, Dixon (2014) contends that although the two prefixes appear superficially similar, they exhibit significant semantic differences: *quasi-X* denotes something that possesses some characteristics of X but is not a full X, while *pseudo-X* refers to something pretending to be like X or resembling X without being X. Consequently, while *quasi(-)* is associated with the lack of a key feature, aligning it with approximation, *pseudo(-)* conveys falseness or imitation (a value defined as *disproximation* by Cappelle, Daus & Hartmann 2023).

1.3. Adopted Approach and Scope of the Paper

The study adopts an onomasiological approach to affix rivalry, focusing on how four different forms vie to express a broader, approximative meaning (many-to-one relationship). Subsequently, the semasiological perspective is also considered, examining the multiple meanings associated with individual prefixoids (one-to-many relationship) (cf. Nagano,

so-called language), *lažipauk* (name for the species *Opiliones*), *lažipčela* ‘fake bee’, *lažisvetac* ‘fake saint’, *lažištípavac* (name for the species *Pseudoscorpiones*), *lažiučenje* ‘fake teaching’, *laži-iluzija* ‘fake illusion’, *laži-mit* ‘fake myth’, and *laži-španjolac* ‘fake Spanish person’. The two adjectival formations are *lažidesni* ‘fake right’ and *lažisvjetski* ‘fake worldwide’.

Bagasheva & Renner 2024). This paper has several objectives. First, grounded in the understanding that base selection has traditionally been viewed as closely tied to the productivity of each affix (Plag 1999; Bauer 2001), it examines the productivity of the four prefixoids. Second, it seeks to examine the differences in collocational preferences of the prefixoids to evaluate the extent of their shared/diverging conceptual content. Third, based on the corpus examples, it aims to provide a description of the semantic values conveyed by each prefixoid. In doing so, the study aims to contribute to the better understanding of the understudied category of morphological approximation in Croatian. Given the lack of prior studies focused on identifying the factors influencing the choice between the examined prefixoids, whether categorically or tendentially, the factors included in this work are based on their presumed relevance and the feasibility of their extraction from the corpus.

The paper is organized as follows. The next subsection (§1.2) briefly reviews previous account on morphological approximation in Croatian. §2 outlines the methodology. In §3, results of the analyses are presented. Finally, §4 provides a discussion of the key findings, draws conclusions and suggest possible directions of future research on this topic.

2. Data and Method

The analysis draws on data from the recently released CLASSLA-web.hr corpus (Ljubešić & Kuzman 2024; Ljubešić, Rupnik & Kuzman 2024), a contemporary Croatian web corpus consisting of nearly 2.2 billion words. This corpus, available for download, is particularly valuable due to its recency, as it primarily comprises data from the .hr internet top-level domain collected in 2021 and 2022, ensuring relevant and up-to-date data for synchronic linguistic research. For each of the four approximative prefixoids under investigation, CLASSLA corpus has been queried for <PREF_{APPRX}> nominal⁴ ngrams as well for nominal formations that start with <PREF_{APPRX}-> and <PREF_{APPRX}>, in order to account for constructions in three orthographic variants, i.e. block use (univerbation) (e.g., *nazoviprijatelj* ‘pseudo-friend’), hyphenated use (e.g., *nazovi-prijatelj*), and separated use (juxtaposition) (e.g., *nazovi prijatelj*). While this procedure yielded high recall for most of the prefixoids,

⁴ A brief corpus search revealed that nominal formations are by far the most frequent. Nonetheless, the analysis of non-nominal formations may be useful in order to assess whether the prefixoids exhibit any categorial constraints or preferences.

manual verification was still required, particularly for *nazovi(-)*. As previously noted, Slavic prefixoids *nadri(-)* and *nazovi(-)* originate from the 2nd person imperative forms of the verbs *nadrijeti* ‘to begin’ and *nazvati* ‘to name, to call’. While *nadrijeti* is extremely rare (with only two attestations in the reference corpus), *nazvati* is quite frequent, with 265,980 occurrences. This presented a significant challenge in extracting relevant instances, as many of the examples of debonded forms of *nazovi(-)* were actually imperative constructions, resulting in numerous false positives, as illustrated in (5):

- (5) *Mislim da ti je jedini način podmiriti dug ili **nazovi banku** i pitaj dali to mogu prolongirati.*

‘I think the only way is to settle the debt or call the bank and ask if they can extend it.’

Additionally, only formations with a base that appears at least 10 times in the corpus were included. This threshold was set after observing that bases with lower frequencies were often either non-existent or orthographically compromised.

The dimensions of the final dataset are illustrated in Table 1.

Tab. 1: Size of the dataset

CLASSLA corpus	Tokens	Types
	21,353	4,270

The analysis is structured in two main steps and comprises a quantitative and qualitative analysis of the data. The data was processed using *R* (version 4.4.1) programming language (R Team 2022). In recent years, quantitative approaches to rivalry have emerged, primarily aimed at investigating the discriminative properties of competing affixes. Consequentially, a range of statistical methods has been employed to assess the impact of (non-)structural factors in resolving rivalry (Huyghe & Varvara 2023; Salvadori, Varvara & Huyghe 2024).

In this study, first, an in-depth analysis of the prefixoids’ productivity was conducted. Various approaches have been proposed for quantitatively evaluating productivity, all operating under the premise that a process is more productive when it generates a greater number of lexemes (Fernández-Domínguez 2013). To assess the vocabulary size of the four prefixoids, i.e. the number of types they form in relation to the increasing number of tokens generated by the process, the study first examines the prefixoids’ vocabulary growth curves (VGC). Furthermore, recognizing the importance of considering multiple variables, rather

than relying solely on a single measure of productivity (Hartmann 2018), three measures of productivity were calculated – Moving-Average Type-Token Ratio (MATTR) (Covington & McFall 2010), Shannon entropy (see, e.g., Hein & Brunner 2020; Evert & Baroni 2022), and Potential Productivity *P* (Baayen 1992, 2009) – each capturing different dimensions of the phenomenon.

MATTR (Covington & McFall 2010) is a sample-size independent variant of the well-known Type-Token Ratio (TTR). MATTR employs a moving window approach to estimate TTR values for each successive window of fixed length. Initially, a window length is chosen, such as 100 words, and the TTR for words 1–100 is computed. The TTR is then calculated for words 2–101, followed by 3–102, and so on, until the end of the dataset. The final score is obtained by averaging all the estimated TTR values. Unlike TTR, MATTR final score is considered unaffected by text length or any statistical assumptions.

In recent years, an information-theoretic approach has been considered and the connection between entropy (Shannon 1948) and productivity in linguistic samples has been observed⁵ (Sundquist 2020). Entropy can be understood as a measure of the uniformity of a probability distribution, or, in other words, a measure of uncertainty. It is calculated from a sample of tokens all derived with a specific morpheme. Initially, we tally the occurrences of each type in the sample, resulting in a type frequency distribution. Next, we convert the type frequency distribution to a probability distribution using maximum likelihood estimation. Finally, we compute the entropy of this probability distribution, measured in bits. Higher entropy indicates that tokens are more evenly distributed among types, which suggests greater productivity. Conversely, low entropy means that most attestations correspond to a few highly frequent types, indicating lower productivity (Barðdal et al. in prep.).

Finally, Potential Productivity *P* (Baayen 1992, 2009) refers to the relation between the number of hapax legomena formed with an analyzed affix in a sufficiently large corpus and the total number of tokens of that affix in the corpus (Gaeta & Ricca 2015). While *P* is a time-honored and widely-applied measure, it suffers from a well-known methodological

⁵ As noted by Gries (2015), entropy is also related to contextual distinctiveness/diversity and learning. For instance, Goldberg et al. (2004) in their work on argument structure generalizations demonstrate that a more skewed distribution (with higher entropy) is learned more effectively than a more balanced one (with lower entropy).

problem – its high sensitivity to sample size. When sample size differs (and it is to be expected that when comparing more morphemes one will be more frequent than another), that difference will affect the productivity measures calculated for those samples, rendering their comparison not interpretable⁶. To obtain a methodologically sound comparison of Potential Productivity values, the token count for individual patterns must be standardized (Gaeta & Ricca 2006). To address this issue, a method known as bootstrapping was applied. From each of the four samples, the same fixed number of tokens was picked randomly. It was determined that a sample size of 1200 would be used, ensuring repeated sampling without significant overlap for the least frequent prefixoid, namely *nadri(-)*, with 1714 tokens. With all groups now standardized to the same size, MATTR, entropy, and Potential Productivity *P* were calculated for each group, and this process was iterated 500 times, each time selecting new random samples with the *with replacement* function. This repetition was crucial for generating more reliable results. By executing this process 500 times, we can ensure a more robust representation of the data, enhancing the validity of the analyses.

After analyzing the morphological productivity of the prefixoids, we shift our focus to a sociolinguistic variable, namely text genre in which the formation appears in, by checking the *genre* tag available in CLASSLA corpus.

Finally, we examine the prefixoids' semantic properties. To examine their distributional/ collocational preferences, Multiple Distinctive Collexeme Analysis (MDCA), a method within the broader framework of collocation analysis (Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004) is applied. The results from MDCA serve as input for a multifactorial analysis technique known as Correspondence Analysis (CA) (Benzécri 1973; Greenacre 2017). The examinations draw on the distributional hypothesis and build upon the assumption that “the degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear”

⁶ That is because *P* is conceptually related to the aforementioned VGCs, as it represents the slope at its end-point, when the maximum number of tokens has been observed. Since the position on the curve is often unknown (for instance, whether we have observed the entire curve or only, for example, the first 20% due to a small sample size), comparing *P* values across divergent sample sizes is problematic (Evert & Lüdeling 2001).

(Lenci 2008: 3)⁷. The principle asserts that a correlation between distributional similarity, i.e. linguistic contexts in which an element is observed, and meaning similarity enables us to infer the latter from the former. The working hypothesis posits that an overlap in collocational preferences among *kvazi(-)*, *nadri(-)*, *nazovi(-)*, and *pseudo(-)* would indicate a shared conceptual content, thus classifying them as near-synonyms. On the other hand, differing collocational patterns would indicate that these prefixoids impose distinct construals. To examine the hypothesized near-synonymy among the prefixoids, MDCA is employed to contrast the four near-synonymous constructions (<*kvazi* + noun>, <*nadri* + noun>, <*nazovi* + noun>, <*pseudo* + noun>). In this analysis, orthographic variations are not deemed relevant, and the input data for each construction is given by the sum of the frequencies of its three orthographic variants. MDCA filters out overlapping collocates and focuses solely on the nouns that are idiosyncratic to each prefixoid, enabling the classification of distinctive nouns based on their function and meaning, thereby providing a deeper understanding of the individual specificities of the four prefixoids. The results from MDCA are crucial for the subsequent Correspondence Analysis (CA). CA, a distance-based clustering technique, visually represents cross-tabulations on a two-dimensional plot, facilitating the mapping of correlations between lexical items. In this study, CA demonstrates how prefixoids imply specific construals based on the nouns they are associated with. The input for CA is derived from the cross-tabulation of frequencies of the 50 most distinctive collexemes of each of the four prefixoids, as identified by MDCA. These analytical methods collectively offer a comprehensive exploration of the relationship between the examined prefixoids, enriching our understanding of their semantic properties.

⁷ When discussing distributional hypothesis, another approach to capturing the semantics of prefixes has to be acknowledged – vector space models (for previous accounts on approximative prefixes, cf. Van Goethem & Norde 2020; Cappelle, Daus & Hartmann 2023). Pre-trained models for Croatian include word embeddings derived from the skip-gram model of fastText (Terčon & Ljubešić 2023), and a transformer model *BERTiĆ* (Ljubešić & Lauc 2021). However, these embeddings are based on multiple corpora: the skip-gram model embeddings originate from *hrWaC* corpus (Ljubešić & Klubička 2016) and the *MaCoCu-hr* corpus (Bañón 2022), while *BERTiĆ* is trained on a combination of ten datasets totaling over 8 billion tokens of written text in Bosnian, Croatian, Montenegrin, and Serbian (closely related languages classified under the same *hbs* identifier (Serbo-Croatian macrolanguage) by the ISO-693-3 standard). Due to time constraints, it was not feasible to train a new model exclusively on the CLASSLA dataset. To maintain methodological consistency with other analyses presented in this study, the exploration of Croatian approximative prefixes using vector space models is deferred to a future study.

As the last step of the semantic analysis, semantic values conveyed by the prefixoids are examined. The initial approach involved analyzing the 50 most distinctive collexemes of each prefixoid, as used for the Correspondence Analysis, under the assumption that these would be semantically representative of each prefixoid. However, as the annotation process progressed, two concerns arose. Firstly, it became apparent that a single lexical item may exhibit diverging meanings depending on its context and the register of use. Although the noun that the prefixoid combines with typically influences its reading, this influence is not always straightforward. For example, in (6), *pseudodokumentarac* ‘pseudo-documentary’ refers to a recognized film genre, devoid of any subjective connotation, whereas in (7) it carries a pejorating sense (further emphasized by the sentence context and the sarcastic portrayal of invited guests), suggesting that the speaker is ridiculing the entire situation.

- (6) *Odlučivši se na znatno kraćenje radnje u odnosu na Shakespeareov predložak, Welles je projekt realizirao pune tri godine [...] o čemu je 1978. napravio **pseudodokumentarac** “Snimajući Otela”.*

‘Having decided on significantly shortening the plot compared to Shakespeare’s template, Welles spent a full three years realizing the project [...], about which, in 1978, he made a pseudo-documentary “Filming Othello”’

- (7) *Složiti će i **pseudodokumentarac** o Novom valu unutar kojeg će probrani sa šalovima oko vrata [...] reći tri prigodne floskule.*

‘He will also put together a pseudo-documentary about the New Wave, in which a selected few with scarves around their necks [...] will say three appropriate clichés.’

For this motive, annotating only the types was deemed insufficient. Additionally, it became clear that the 50 most distinctive collexemes did not adequately capture the full semantic spectrum of each prefixoid, as the range of meanings identified was more limited than the one observed in a random sample of derivatives. To provide a more comprehensive overview, a different approach was adopted. Given the large number of relevant tokens (21,353), annotating all tokens according to their specific semantic value was not feasible. Therefore, we opted to semantically annotate a random sample of 500 tokens per prefixoid, with the expectation that this sample size would be sufficient to represent the range of meanings each prefixoid can convey (cf., a.o., Masini & Micheli (2020) who annotated a sample of 219 tokens, and Micheli (2023) who worked with samples of 100 tokens).

3. Results

3.1. Morphological Productivity

Table 2 summarizes key characteristics of the analyzed data, including the number of tokens, types, and hapax legomena. The frequencies of nominal formations exhibit variability, with token counts ranging from 1,714 to 8,647 and type counts spanning from 277 to 2,533.

Type count represents one of the most straightforward measures of morphological productivity – the greater the number of types a morpheme generates, the more productive it is considered (Bauer 2001; Zeldes 2012) – and is seen as an estimation of the ‘extent of use’ or the ‘profitability’ of a morphological category (Corbin 1987). An examination of the results in Table 2 reveals that *kvazi(-)* forms the most types (2,533), followed by *nazovi(-)* (1,498), and *pseudo(-)* (1,446), while *nadri(-)* generates the fewest types, viz. 277, 10.94 times fewer than *kvazi(-)*. Consequentially, *prima facie*, *kvazi(-)* appears to be the most productive prefixoid, whereas *nadri(-)* seems the least productive one. However, due to differences in sample sizes across prefixoids, comparing raw type counts is not meaningful.

Tab. 2: Statistical overview of prefix usage

PREFIXOID	Tokens	Types	Hapax legomena
<i>kvazi(-)</i>	8,647	2,533	1,557
<i>nadri(-)</i>	1,714	277	189
<i>nazovi(-)</i>	3,230	1,498	980
<i>pseudo(-)</i>	7,942	1,446	816

In addition to analyzing the number of types formed by each prefixoid, an initial understanding of their productivity can also be gained through the analysis of their vocabulary growth curves (VGCs). VGCs show the vocabulary size, i.e. the number of types, in relation to the increasing number of tokens generated by the examined four processes. Typically, a VGC displays a characteristic shape because the number of types observed for a given number of tokens associated with a particular prefixoid is a monotonically increasing function of the number of tokens: as more tokens are sampled, more types are observed. Initially, the curve rises fairly steeply, but, as more tokens are encountered, the rate at which new types are observed decreases, causing the growth of

the curve to slow down. For morphological categories with a finite number of types, the curve eventually plateaus, indicating that no new types are observed beyond a certain token count. A relatively unproductive process typically displays a shallow or asymptotic VGC, where vocabulary growth stabilizes early, reflecting fewer instances of novel types (Baayen 2001; Evert & Lüdeling 2001).

Figure 1 illustrates the VGCs for the four examined prefixoids. Alongside the observed vocabulary curve, depicted as a dashed line, the plot also shows an interpolated growth curve. An empirical growth curve often appears irregular due to variations arising from the non-random distribution of words in a corpus (Baroni & Evert 2014). To achieve a smoother curve, binomial interpolation is used (Baayen 2001). Binomial interpolation utilizes a frequency spectrum to generate expected values of vocabulary size for a given sample size. The interpolated curve obtained using the *vgc.interp* function from *zipfR* (Evert & Baroni 2022) is presented as a solid line. The interpolated curve was evaluated using Root Mean Square Error (*RMSE*) and R^2 . The R^2 values are high across all prefixoids, ranging from 0.993 to 0.999, while the *RMSE* values span from 6.070 to 14.684. Both values confirm that the interpolated curves are performing as desired.

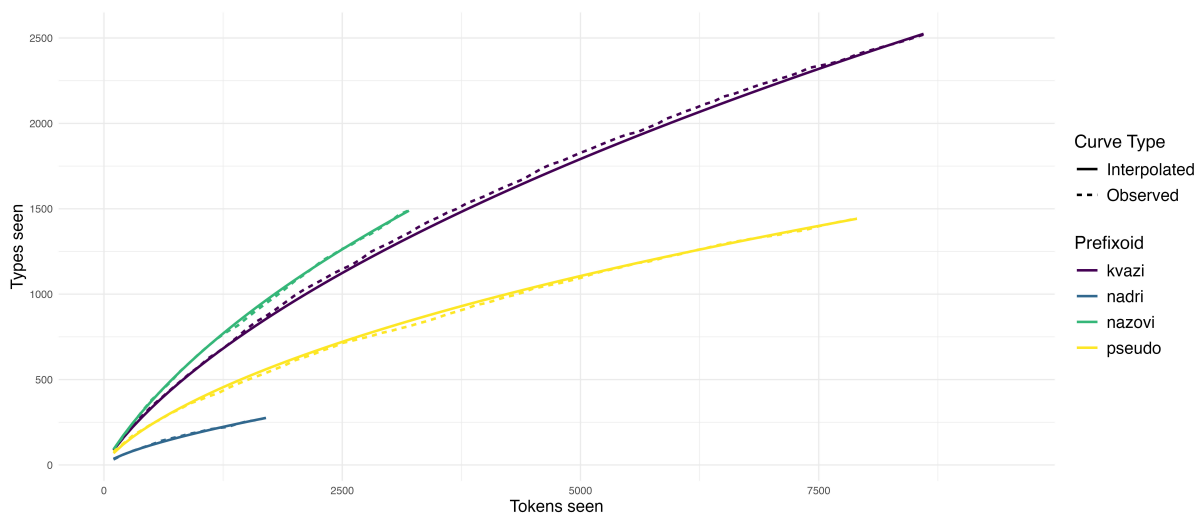


Fig. 1: Vocabulary growth curves for the four prefixoids

The VGCs for the four prefixoids reveal distinct patterns. First, the plot clearly adheres to the previously described general trend: most prefixoids exhibit a rapid initial growth in the number of types, which gradually slows as the number of tokens increases. The prefixoid *nazovi(-)* displays the steepest initial rise, indicating a quicker discovery of new types. On

the other hand, *kvazi(-)* and *pseudo(-)* demonstrate a more moderate growth trajectory, while *nadri(-)* exhibits the flattest curve, indicative of its limited capacity to generate novel types.

Following these initial observations of productivity, a more detailed analysis was conducted using the three aforementioned productivity measures: Moving-Average Type-Token Ratio (MATTR), entropy, and Potential Productivity P . These measures were selected for their ability to capture different aspects of morphological productivity. While MATTR measures balance in usage, Potential Productivity estimates the likelihood of encountering a new type, and entropy can be seen as a measure of uncertainty (unpredictability) in the type-frequency distribution. To mitigate the impact of varying sample sizes, a standardized random sample of 1,200 tokens (70% of the least frequent prefixoid's token count, in order to allow repeated sampling with minimal overlap) was selected for each prefixoid. For each sample, MATTR (100-word window), entropy and Potential Productivity P were computed. Random sampling was repeated 500 times, with all three measures calculated for each iteration. The fluctuations in the values of these three productivity measures across the 500 randomly sampled 1,200-token sets are depicted in Figure 2.

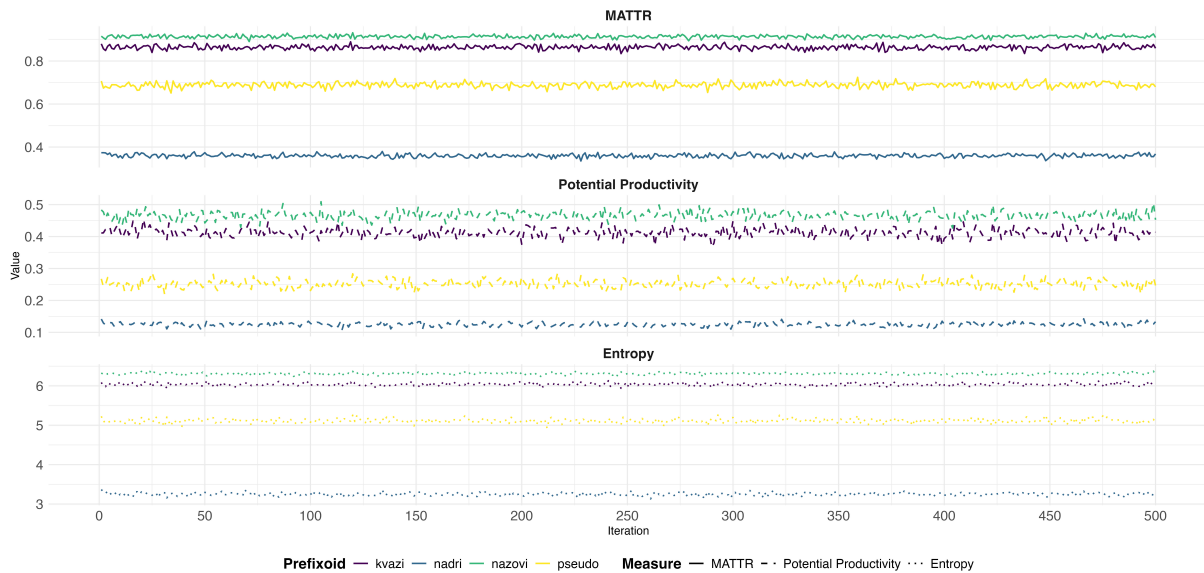


Fig. 2: Fluctuations of the productivity measures values over 500 fixed-size samples

The plot reveals that fluctuations are most pronounced for Potential Productivity values, moderately pronounced for MATTR, and least pronounced for entropy. In fact, min-max

normalized Coefficient of Variation across the three measures for all prefixoids amounts to 0 for entropy, 0.152–0.277 for MATTR, and 1 for Potential Productivity, indicating it as the most volatile measure, very sensitive to changes in data composition. These findings validate the chosen method, i.e. bootstrapping with 500 fixed-size samples, while also highlighting the risk of relying on a single random sample, which could result in an “extreme” value rather than an average (representative) one. To facilitate a meaningful comparison of productivity across the prefixoids, Figure 3 presents the MATTR, Potential Productivity, and entropy results, with the median values displayed for each measure and each prefixoid⁸. This approach offers a balanced perspective, mitigating the influence of sample variability and providing a clearer insight into the relative productivity dynamics of each prefixoid.

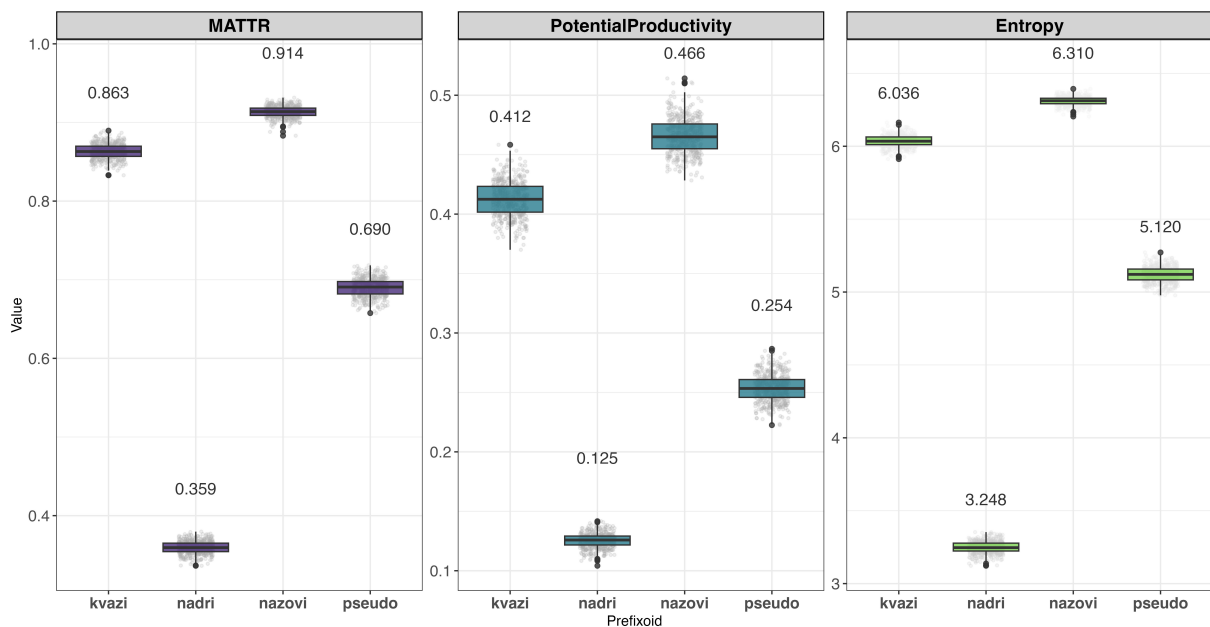


Fig. 3: MATTR, Potential Productivity, and entropy box plots including median values

The results of the fixed-size sample analysis are consistent and the rankings according to three productivity measures align perfectly (Pearson correlation coefficient spans from 0.967

⁸ In order to statistically confirm that the observed differences between productivity measures were significant, since the assumption of homogeneity of variances was violated, the Kruskal-Wallis test was performed. The test revealed significant differences in entropy ($\chi^2 = 1874.1$, $p < 2.2e-16$), MATTR ($\chi^2 = 1874.1$, $p < 2.2e-16$), and Potential Productivity ($\chi^2 = 1869.3$, $p < 2.2e-16$) across the four prefixes. Dunn’s post-hoc tests with Bonferroni correction further confirmed that all pairwise comparisons between the prefixes were statistically significant ($p < 0.05$). This step ensured that the trends seen in the boxplots are statistically robust, thereby providing a more rigorous basis for interpreting the productivity differences among the prefixes.

for Potential Productivity–entropy correlation to 0.999 for MATTR–entropy correlation). This suggests that, while the adopted measures capture different aspects of productivity, when the fixed-sample size rule is applied, they seem, in fact, highly correlated and can be used interchangeably for comparative purposes. According to all three measures, *nazovi(-)* is the most productive prefixoid, closely followed by *kvazi(-)*. *Pseudo(-)* ranks third in productivity, while *nadri(-)* is deemed the least productive, with a notable gap separating it from the other prefixoids. The findings corroborate the well-established postulate that rival affixes typically exhibit differences in productivity levels (cf. Bybee 1985; Baayen & Lieber 1991; Plag 1999; Bauer 2001; Gaeta & Ricca 2015).

3.2. Prefixoids Across Text Genres

Apart from various linguistic factors, competition between rival forms can also be influenced by the register (genre) in which they appear, with specific affixes being favored in certain contexts. To examine the relationship between text genre and prefixoid selection, this study makes use of the genre annotations present in the CLASSLA corpus. The corpus was automatically annotated for genres using the Transformer-based X-GENRE classifier (Kuzman, Mozetič & Ljubešić 2023) and the following genre categories are used: News, Information/Explanation, Promotion, Opinion/Argumentation, Instruction, Legal, Prose/Lyrical, Forum, Other and Mix⁹.

To evaluate the dependency between the choice of a prefixoid and the genre it appears in, a Pearson's χ^2 test was conducted. The test was performed on a contingency table that cross-tabulates the frequencies of four Croatian prefixoids across various genres. The test results ($\chi^2 = 5239.2$, $df = 30$, $p\text{-value} < 2.2e-16$) indicate a statistically significant relationship between the prefixoid used and the genre of the text. This significant p-value highlights a strong association between genre and prefixoid choice, suggesting that different genres may prefer distinct prefixoids.

The bar chart in Figure 4 illustrates the distribution of four Croatian prefixoids across the ten aforementioned genres. The values are standardized within each genre to show the

⁹ Description and examples of each genre category are available at <https://huggingface.co/classla/xlm-roberta-base-multilingual-text-genre-classifier> (accessed 24 August 2024).

relative frequency of each prefixoid as a percentage of the total number of prefixoids within that specific genre.

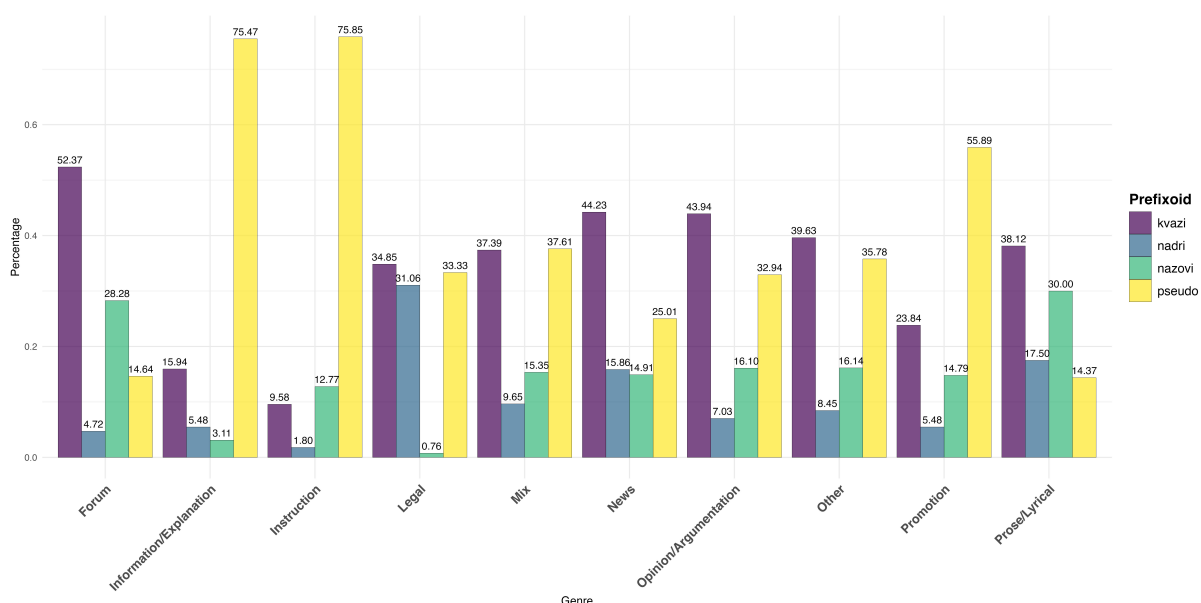


Fig. 4: The standardized frequency of the prefixoids across genres

Observing the bar chart, it is visible that:

kvazi(-) has prominent appearances in genres like Forum (52.4%) and News (44.2%), while it is least prevalent in Instruction (9.58%) and Information/Explanation (15.9%);

nadri(-) presents a notable peak in Legal (31.1%), while it is the least frequent in Instruction (1.80%) and Forum (4.72%);

nazovi(-) exhibits a relatively even presence across all genres with the highest frequencies in Prose/Lyrical (30%) and Forum (28.3%), while only one token (0.76%) formed with *nazovi* is found in Legal;

pseudo(-) is the prominent in Instruction (75.8%) and Information/Explanation (75.5%). It is the least dominant in Prose/Lyrical (14.4%) and Forum (14.6%).

Several compelling observations can be drawn from the genre distribution of the prefixoids. First, a near-perfect complementarity is evident between *nazovi*(-) and *pseudo*(-). The genres in which *nazovi*(-) is the most frequent (Prose/Lyrical and Forum), are in fact genres in which *pseudo*(-) is the least dominant. The complementarity observed between the two elements suggests a form of genre-based niche differentiation that reduces direct competition. This genre-based divergence could likely parallel their distinct collocational preferences, as will be showed in the Correspondence Analysis (Figures 5 & 6),

where *nazovi(-)* and *pseudo(-)* are located on the very opposite sides of the biplot, reinforcing their functional and semantic dissimilarity. On the other hand, *kvazi(-)* and *nazovi(-)* display a more aligned genre distribution, particularly in genres like Forum and Prose/Lyrical, in which they are the most frequent prefixoids. This suggests that *kvazi(-)* and *nazovi(-)* may have a closer functional or semantic relationship, which also justifies their overlap in the Correspondence Analysis (Figures 5 & 6). Furthermore, it is curious to observe how the two prefixoids of Slavic-origin exhibit distinct genre-specific usage patterns, especially in Legal. While *nadri(-)* observes its peak in this genre (primarily due to the strong presence of modified profession denoting nouns such as *nadriliječnik* and *nadripisar*), at the same time just a single occurrence formed with *nazovi(-)* is found in Legal, representing the lowest share (0.76%) of any prefixoid across all ten genre categories. These findings highlight genre as a significant factor influencing the selection between rival prefixoids, demonstrating that different communicative contexts, i.e. less-related semantic fields, distinctly shape morphological choices. This underscores the importance of genre-specific factors in affix rivalry, revealing how certain genres favor specific prefixoids while disfavoring others, thereby influencing the functional landscape of morphological approximation.

3.3. Collocational Behavior

Having examined the productivity and genre distribution of the four prefixoids, our focus shifts to their semantics. Rival approximative affixes, even when used in analogous contexts, often exhibit distinct distributional tendencies. To analyze the extent of overlap in collocational preferences between the four prefixoids, first we identify the number of nominal bases they share. Subsequently, we employ Multiple Distinctive Collexeme Analysis (MDCA) to extract the most distinct collexemes of each prefixoids. This approach isolates the bases that are particularly characteristic of each prefixoid, highlighting their unique collocational patterns. We conclude the semantic analysis by visualizing the most distinct collexemes by means of Correspondence analysis (CA). The combined use of MDCA and CA offers a comprehensive understanding of how these rival prefixoids interact within the semantic landscape, providing a clearer picture of their distinct yet overlapping roles.

3.3.1. Shared Bases

Table 3 presents the number of nominal bases shared by each pair of prefixoids.

Tab. 3: Number of shared nominal bases by pairs of prefixoids

PREFIXOID	<i>kvazi(-)</i>	<i>nadri(-)</i>	<i>nazovi(-)</i>	<i>pseudo(-)</i>
<i>kvazi(-)</i>		163	658	586
<i>nadri(-)</i>			124	97
<i>nazovi(-)</i>				341
<i>pseudo(-)</i>				

As observed by comparing the values in the Table 3, *kvazi(-)* and *nazovi(-)* share the highest number of bases (658), suggesting they might be more closely related to each other than with other prefixes. In contrast, *nadri(-)* shares the fewest bases with the other prefixoids, indicating a more distinct usage pattern¹⁰. While the number of shared bases offers some insight into the semantic and functional relatedness of the four prefixoids, it provides only a partial view, and ulterior analyses are required in order to obtain a more complete picture of the overall relationship and the extent of the semantic overlap.

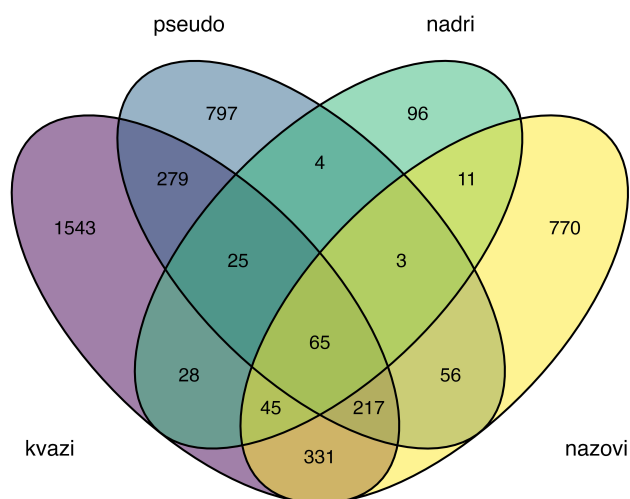


Fig. 5: The overlap of base nouns between the prefixoids

¹⁰ To quantify the degree of overlap between pairs of prefixoids more precisely, Jaccard Similarity Indices were calculated. The Jaccard Similarity Index (JSI) measures similarity between finite sample sets and is defined as the size of the intersection divided by the size of the union of the sample sets. The Index ranges from 0 (no overlap) to 1 (complete overlap). As anticipated, the highest JSI was observed between *kvazi(-)* and *nazovi(-)* (0.195), followed by *kvazi(-)* and *pseudo(-)* (0.173), and *nazovi(-)* and *pseudo(-)* (0.131). The lowest JSI was noted between *nadri(-)* and *pseudo(-)* (0.059), suggesting minimal shared bases between these two prefixoids.

Figure 5 presents a four-way Venn diagram that illustrates the shared and unique bases among the four examined prefixoids. Unlike the simple pairwise comparisons shown in Table 3, this diagram provides a detailed visualization of the specific overlap counts between different sets without aggregating them. Each ellipse represents the set of bases associated with a specific prefixoid; for instance, the light blue ellipse represents *kvazi*'s bases. Numbers within the non-overlapping sections indicate bases unique to each prefixoid, such as the 1,543 bases that *kvazi*(-) does not share with the other prefixoids. Conversely, numbers in overlapping sections indicate shared bases, such as the 331 bases shared exclusively between *kvazi*(-) and *nazovi*(-). The central number (65) represents the bases common to all four prefixoids. To determine the total shared bases between two prefixoids, such as *kvazi*(-) and *nazovi*(-), one must sum all overlapping regions between their respective ellipses, including those overlapping with other ellipses. Specifically, *kvazi*(-) and *nazovi*(-) share 658 bases in total: 331 bases shared exclusively between the two, 217 bases shared with *pseudo*(-), 45 bases shared with *nadri*(-), and, finally, 65 bases shared by all four prefixoids. This visualization highlights the degree of overlap between the prefixoids, providing insight into their relational dynamics based on shared bases. For instance, *kvazi*'s greater base overlap with *nazovi*(-) and *pseudo*(-) than with *nadri*(-) could suggest a closer functional or semantic connection between these prefixes.

Having established the proportion of shared bases, it is relevant to further examine the 65 nominal bases common to all four prefixoids. Notably, 38 of these shared bases are directly related to individuals, either as profession-denoting nouns (e.g., *filozof* 'philosopher', *novinar* 'journalist') or as cultural/political terms (*roker* 'rocker', *ljevičar* 'leftist'). Additionally, terms such as *demokracija* 'democracy', *komunizam* 'communism', and *institucija* 'institution' are found, reflecting the frequent use of approximation prefixoids in contexts of societal, and particularly political, discourse. What can be observed in these groups, just like in examples (1)–(4), is that the prefixoids exhibit rival behavior and semantic equivalence (at least at a coarse-grained level), with all four signaling a person, object, or concept that is not a typical exemplar of the base noun, often with reference to inadequate performance. While these (minimal) pairs might be useful for highlighting differences between rival prefixoids by neutralizing base differences, they may not capture all distinctions between rival processes, as certain discriminative properties of the base may prevent the formation

of competing lexemes (cf. Huyghe & Varvara 2023). For this motive, we leave the study of formations with identical bases for future research and instead focus on the identification of collexemes unique to each approximative construction, believing this approach offers deeper insight into the semantics of the prefixoids.

3.3.2. Multiple Distinctive Collexeme Analysis

The data concerning the quantity of shared bases presented in the previous section made us hypothesize about the nature of the semantic relationship between the four prefixoids. While some prefixoids share more bases (hence, could be semantically more similar), others do not seem to share many bases with other prefixoids, indicating a potentially more peculiar collocational behavior. A possibility of contrasting more related, near-synonymous constructions in their respective synchronic collocational preferences is made possible by Multiple Distinctive Collexeme Analysis (MDCA) (Gries & Stefanowitsch 2004; Stefanowitsch 2013). MDCA enables the identification of collexemes that are unique to specific constructions, moving beyond mere raw frequency by abstracting common elements and focusing on distinct usage patterns. By systematically examining usage-based, pattern-specific properties through statistical analysis, MDCA assesses asymmetries in the relative frequencies of co-occurring lexical items (Stefanowitsch & Flach 2020), highlighting collexemes that occur significantly more frequently with one construction over another. The input required for MDCA is a data frame, which can be formatted as either a raw frequency list (one observation per line) or an aggregated frequency list that includes a third column for the construction's frequency. The MDCA script used in this study is based on Flach's (2021) *collex.covar* function from the package *constructions*. Furthermore, following the approach outlined by Proisl (2022), units of analysis (corpus size) consist of all (722,422,618) nominal tokens in the CLASSLA corpus. Finally, association – or to be more precise, a combination of frequency, association, and dispersion (cf. Gries 2019, 2022) – is measured using the log-likelihood ratio (G^2), “the most frequently used [association] measure”¹¹ (Gries 2019:

¹¹ Recent discussions have focused on using measures like residuals of chi-squared to further reduce computational costs. Additionally, there is an ongoing debate on calculating more than one association measure value to address the issue of conflating frequency, mutual/unidirectional association, and dispersion into one measure (as G^2 does) (Gries 2019, 2023; Liao, Gries & Wulff 2024). However, given that the primary objective of this analysis is exploratory/descriptive, it is essential to recognize that the “conflation of frequency and association makes for a good exploratory tool” (Liao, Gries & Wulff 2024: 13), and “an

150). The degree of attraction between the construction and the collexeme (and vice versa, since G^2 is a bidirectional measure) is very significant at (at least) $p < 0.0001$.

Tables 4–7 present the ten most distinctive collexemes for each of the four analyzed prefixoids. OBS stands for *observed frequency*, indicating how often each prefixoid-noun pairing appears in the corpus. EXP represents the *expected frequency*, i.e. anticipated by chance under the null hypothesis. COLL.STR.LOGL stands for *collostructional strength log-likelihood*, a measure of the association between the two slots of the construction.

Tab. 4: The 10 most distinctive nominal collexemes of *nadri(-)*

<i>nadri(-)</i>	OBS	EXP	COLL.STR.LOGL	$\Delta P1$	$\Delta P2$
<i>liječništvo</i> ‘medical profession’	453	35.2	2435.84	0.2625	0.9415
<i>liječnik</i> ‘doctor’	393	32.2	1934.40	0.2267	0.8881
<i>pisarstvo</i> ‘scribal profession’	141	11.0	731.27	0.0817	0.9282
<i>pisar</i> ‘scribe’	72	72	370.64	0.0417	0.9253
<i>ljekarstvo</i> ‘medicine’	13	1.0	66.50	0.0075	0.9228
<i>obrtnik</i> ‘craftsman’	12	0.9	61.38	0.0070	0.9228
<i>veterinarstvo</i> ‘veterinary medicine’	9	0.7	46.02	0.0052	0.9226
<i>liječenje</i> ‘treatment’	10	0.9	40.65	0.0057	0.7560
<i>obrt</i> ‘craft/trade’	6	0.5	30.67	0.0035	0.9225
<i>majstor</i> ‘handyman’	17	4.7	21.88	0.0077	0.2015

Tab. 5: The 10 most distinctive nominal collexemes of *kvazi(-)*

<i>kvazi(-)</i>	OBS	EXP	COLL.STR.LOGL	$\Delta P1$	$\Delta P2$
<i>parcijala</i> ‘repetition of some university courses’	84	33.2	156.60	0.0096	0.60745
<i>menadžer</i> ‘manager’	127	56.5	152.82	0.0133	0.4965
<i>kristal</i> ‘crystal’	78	30.8	145.39	0.0089	0.6073
<i>renta</i> ‘rent’	54	21.30	100.56	0.0062	0.6066
<i>čestica</i> ‘particle’	50	19.70	93.09	0.0057	0.6065
<i>novinar</i> ‘journalist’	198	118.5	87.04	0.0150	0.2688
<i>navijač</i> ‘fan’	67	30.4	75.35	0.0069	0.4769
<i>političar</i> ‘politician’	147	84.9	73.91	0.0117	0.2917
<i>grupa</i> ‘group’	47	19.3	72.79	0.0052	0.5656
<i>intelektualac</i> ‘intellectual’	158	94.4	69.72	0.0119	0.2691

association measure combining (a lot of) frequency and (a little bit of) association is still a good option” as it provides a “heuristically useful amalgam of two kinds of information” (Gries 2023: 372).

Tab. 6: The 10 most distinctive nominal collexemes of *nazovi(-)*

<i>nazovi(-)</i>	OBS	EXP	COLL.STR.LOGL	$\Delta P1$	$\Delta P2$
<i>prijatelj</i> ‘friend’	48	11.6	92.14	0.0121	0.5066
<i>album</i> ‘album’	20	3.4	65.27	0.0055	0.7914
<i>sud</i> ‘court’	24	5.3	52.06	0.0062	0.5664
<i>bog</i> ‘god’	18	3.7	43.33	0.0047	0.6215
<i>banka</i> ‘bank’	12	2.1	37.06	0.0033	0.7618
<i>hrvat</i> ‘Croat’	27	8.6	34.21	0.0061	0.3485
<i>grad</i> ‘city’	14	3.01	32.89	0.0036	0.5756
<i>pjesma</i> ‘song’	10	1.8	27.80	0.0030	0.6837
<i>sloboda</i> ‘freedom’	17	4.7	26.42	0.0045	0.3989
<i>komentar</i> ‘comment’	10	2.3	20.49	0.0028	0.5170

Tab. 7: The 10 most distinctive nominal collexemes of *pseudo(-)*

<i>pseudo(-)</i>	OBS	EXP	COLL.STR.LOGL	$\Delta P1$	$\Delta P2$
<i>znanost</i> ‘science’	1014	397.5	1649.99	0.1197	0.5964
<i>žitarica</i> ‘cereal’	365	134.2	721.96	0.0448	0.6395
<i>kod</i> ‘code’	236	86.3	479.27	0.0291	0.6412
<i>cista</i> ‘cyst’	152	55.6	307.66	0.0187	0.6387
<i>stvarnost</i> ‘reality’	167	63.3	291.47	0.0201	0.6044
<i>jezik</i> ‘language’	91	33.6	173.62	0.0111	0.6261
<i>gen</i> ‘gene’	66	24.1	133.14	0.0081	0.6363
<i>gravidnost</i> ‘pregnancy’	66	24.1	133.14	0.0081	0.6362
<i>stablo</i> ‘tree’	65	23.8	131.12	0.0080	0.6362
<i>dogadaj</i> ‘event’	70	26.06	118.92	0.0084	0.5952

Tables 4–7 provide additional information, specifically the ΔP value (Ellis 2007; Ellis & Ferreira-Junior 2009), which addresses the limitations of G^2 related to its bidirectionality. Unlike G^2 , ΔP is unidirectional (asymmetric), meaning it does not conflate $p(\text{word2}|\text{word1})$ and $p(\text{word1}|\text{word2})$ into a single value. This distinction allows ΔP to identify cases where collexeme 1 strongly attracts collexeme 2, but not vice versa. Additionally, ΔP reflects association independently of frequency, meaning changes in corpus size do not affect the association value. ΔP is divided into two values: $\Delta P1$ measures the predictiveness of the collexeme (slot 2) for the construction (slot 1), whereas $\Delta P2$ quantifies the predictive capacity of the construction (slot 1) for the collexeme (slot 2) (Gries & Ellis 2015; Gries 2019). As anticipated, an analysis of $\langle \text{PREFIX}_{\text{APPRX}} + \text{noun} \rangle$ constructions reveals that the constructions are more predictive of the noun than the other way around, as indicated by significantly higher $\Delta P2$ values compared to $\Delta P1$. Among the four prefixoids, *nadri(-)* stands out as the

most predictive when considering its ten most distinctive collexemes, with a very high mean $\Delta P2$ of 0.8331. The highest predictiveness is observed with *nadriliječništvo* ‘quackery’, where the construction almost impeccably anticipates its collexeme ($\Delta P2 = 0.94$).

When overlapping collexemes (nouns correlated with two or more analyzed prefixoids) are filtered out by MDCA, the distinct collocational patterns of the prefixoids become more pronounced. In fact, Aronoff’s (2019) *habitat niche differentiation* clearly emerges for *nadri(-)*: nine out of the ten most preferred collexemes denote a profession or a person exercising a profession, while the remaining non-profession noun, viz. *liječenje* ‘treatment’, denotes a job done by the aforementioned medical professional. Moreover, differences in the level of formality of the preferred collexemes are observed. Whereas *nadri(-)* and, especially, *pseudo(-)* combine with nouns that belong to a more formal and sometimes scientific register, *nazovi(-)* and *kvazi(-)* seem to combine more felicitously with nouns from general lexicon.

While MDCA has offered valuable insights, it is important to acknowledge its limitations, especially due to the small number of displayed collexemes for each prefixoid. On the other hand, as noted by Desagulier (2014), increasing the number of collexemes can make it harder to draw meaningful generalizations from the data. Therefore, rather than relying solely on comparing MDCA output tables, it is advisable to use a technique that allows for visualizing the relationships between (a) prefixoids, (b) nouns, as well as (c) prefixoids *and* nouns, by converting the initial matrix into a low-dimensional space. In line with Desagulier’s (2014, 2015) approach, and as already applied in Lacić (2024a), this study will use the output of MDCA as input for Correspondence Analysis (CA).

3.3.3. Correspondence Analysis

Correspondence Analysis (CA) is a multifactorial exploratory statistical technique used to explore relationships and patterns within categorical data (Benzécri 1973; Greenacre 2017). In CA, rows and columns of a contingency table are represented as points in Euclidean space, with their proximity indicating the strength of association. The χ^2 distance, multivariate statistical distance measure akin to Euclidean distance but weighted

by the inverse of the average row profile, measures differences between profiles, positioning rows and columns with similar counts closer together¹².

CA was conducted using the 50 most distinctive collexemes of each of the four prefixoids and the raw frequency of each construction as input. The threshold of 50 most distinctive collexemes was decided in order to maintain the degree of attraction between the construction and the collexeme statistically significant – all included collexemes have log-likelihood significant at (at least) $p < 0.001$. Furthermore, the hypothesis of independence regarding the input data can be rejected, with $\chi^2 = 18093.73$; $p\text{-value} = 0$. In addition, Cramér's V of 0.790 indicates a significant association between the rows and the columns, supporting the notion of a meaningful relationship between the prefixoids and nouns they combine with. CA uses the input frequencies to juxtapose (a) line profiles, i.e. distinctive collexemes (nouns); (b) column profiles, i.e. prefixoids; (c) line profiles *and* column profiles, i.e. nouns *and* prefixoids. The *FactoMineR* package was employed to run CA. Figures 6 and 7 (with collexemes translated in English) display the output of CA.

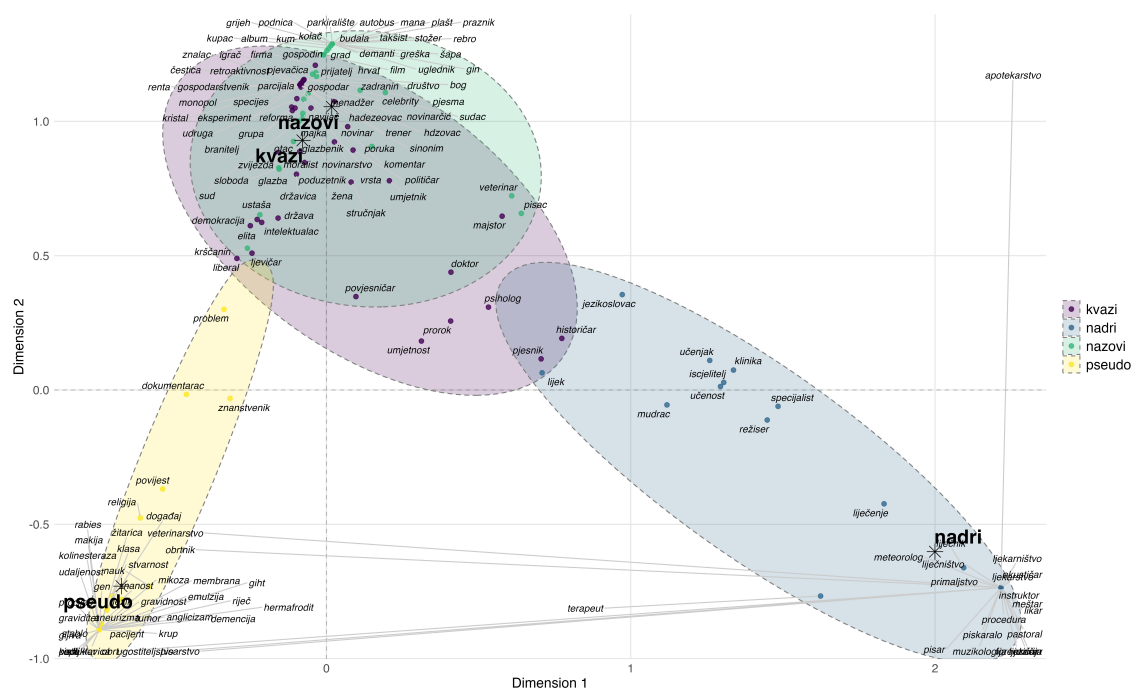


Fig. 6: CA biplot of the <PREF_{APPRX} + noun> construction in CLASSLA corpus

¹² However, interpreting the proximity between rows and columns should be done cautiously, as there is no direct interpretation of row-to-column or column-to-row distances (Levshina 2015).



The CA biplot is constructed using two principal axes of inertia, which intersect to define the average profile of all points in the data cloud. The technique decomposes the overall inertia (Φ^2) – weighted (co)variance obtained dividing the χ^2 statistic by the total sample size – by identifying representative dimensions that condense as much information as possible within each axis corresponding to a dimension. Typically, for reasons of practicality, a plot displays only two dimensions, selected based on their eigenvalues, which measure the amount of information (variation) present along each axis (Levshina 2015; Greenacre 2017). In this analysis, the first axis (dimension 1) represents 45.98% of Φ^2 , while the second axis (dimension 2) represents 37.81% of Φ^2 . There is also a third dimension with an eigenvalue of 16.21%, which is not included in the biplot. Whilst including the third dimension with means of an interactive, tri-dimensional plot would provide additional information regarding the relationships between the analyzed variables, the first two dimensions already account for 83.79% of the variation contained in the input table, allowing for a sufficiently accurate interpretation of the results.

We can start examining the plot and how it juxtaposes four prefixoids by contrasting the two main dimensions. Dimension 1 (horizontal axis), primarily contrasts *pseudo(-)* and *nadri(-)*. Positioned distinctly in the lower left quadrant, *pseudo(-)* occupies a unique space

linked with academic, scientific, or technical terminology that often involves imitation aspects (e.g., *psuedoznanost* ‘pseudoscience’, *pseudodemencija* ‘pseudodementia’, *pseduocista* ‘pseudocyst’). This prefixoid diverges from the societal roles more characteristic of *kvazi(-)* and *nazovi(-)* and instead aligns with terminology that carries an inherent critique of validity or authenticity, particularly in specialized/technical domains. Conversely, on the lower right side, where *nadri(-)* dominates, the focus shifts towards contexts related to medicine, expertise, and professional roles that involve deception or unauthorized practice (e.g., *nadriliječnik* ‘quack’, *nadri-iscjelitelj* ‘quack healer’, *nadriliječenje* ‘quack treatment’). Its unique placement underscores its more specific and legally loaded use, setting it apart from the broader, more socially nuanced applications of the other prefixoids. This isolation reflects its primary association with professional deceit rather than broader conceptualizations of inauthenticity or approximation. In contrast, observing Dimension 2 (vertical axis), a separation of *pseudo(-)* and *nadri(-)* from *kvazi(-)* and *nazovi(-)* is evident. The separation could potentially be interpreted through the lens of the semantic roles these prefixoids play. In fact, as seen from observing the preferred collexemes, *pseudo(-)* attaches to terms in academic, scientific, and technical domains, while *kvazi(-)* and *nazovi(-)* are more common in colloquial or socially-oriented language (e.g., *kvaziprijatelj* ‘pseudo-friend’, *nazovipolitičar* ‘pseudo-politician’) that foster more subjective interpretation. The vertical axis, therefore, might reflect a contrast between formal, domain-specific language (right side) and more general, socially embedded language (left side), which would align with the general theory that formal registers and specialized vocabulary often cluster together due to their shared contexts.

When it comes to clusterings, a distinct cluster can be observed formed with *kvazi(-)* and *nazovi(-)* and their collexemes. Clusters are visually enhanced by confidence ellipses around the groups of data points that are associated with each prefixoid. When ellipses overlap, like in the case of *kvazi(-)* and *nazovi(-)*, it indicates that the base words associated with those prefixoids are similar or used in similar contexts, suggesting a potential functional or semantic similarity between the prefixoids. Furthermore, the size of each ellipse provides insight into the variability or spread of the base words for each prefixoid. Larger ellipses indicate greater variability in how the prefixoid is used with different base words. Returning to the clusters, the distinct position of *pseudo(-)* with scientific and

medical terms corroborates the prefixoid's specialized role in denoting false or imitative phenomena within technical and academic discourses. Conversely, the clear separation of *nadri(-)* suggests another specialized usage, namely denoting fraudulent or unqualified individuals in professional roles.

Following this analysis, the study will proceed with a semantic annotation of selected examples to further elucidate the nuanced meanings of each prefixoid.

3.4. Semantic Annotation

As outlined in §2, the semantic annotation process is based on a 500-token random sample for each prefixoid. The annotation was done manually by one annotator. Notably, the identified semantic values align substantially with those documented by Masini & Micheli (2020) for the Italian *simil-* and Vassiliadou et al. (2023) for *pseudo-* in French, indicating cross-linguistic parallels in how approximation is manifested. To avoid contributing to the already present terminological imbroglio, the labels adopted here partially adhere to the established terminology from the aforementioned studies. In what follows, we first detail the seven identified semantic values, each of which represents a specific manifestation within the broader category of approximation¹³. Subsequently, for each prefixoid, we present the semantic values identified within the 500-token sample. With Y being the output of the prefixoid modification process, and X the nominal base to which the prefixoids apply, the identified semantic values can be defined as follows:

“CLOSE-TO” EVALUATION (CTE): This category captures instances where Y is coming close the state of X, embodying an almost-but-not-quite identity with X. This formation can often be paraphrased using adverbs such as *gotovo* ‘almost’ and *skoro* ‘nearly’, emphasizing a degree-based proximity.

FAKENESS (F): In this value, Y deliberately deceives by pretending to be X without authentically being X. The element of intentional deceit is central, aligning with scenarios where imitation serves to mislead observers into believing Y is genuinely X.

¹³ It should be noted that, from a functional perspective, not all values are approximating at the same level. While concepts such as SUBJECTIVE DEPRECIATIVE EVALUATION or VAGUENESS can be viewed as more closely aligned with approximation, FAKENESS or KIN-CATEGORIZATION are more related to category creation (Masini & Micheli 2020). Given space constraints, we leave a deeper exploration of this complex issue for future research.

ILLEGALITY (ILL): This value is particularly relevant in contexts involving unauthorized or unlawful actions. Here, Y represents an individual or activity that operates outside the boundaries of legal norms.

IMITATION (IM): Involves Y replicating or reproducing aspects of X without the intent of being mistaken for X. Unlike FAKENESS, the absence of intentional deceit distinguishes this category, allowing for a less pejorative reading where imitation is acknowledged but not condemned.

KIN-CATEGORIZATION (K-C): This value identifies Y as an entity closely related to X but distinct in its defining characteristics. Y's identity is rooted in X but deviates enough to form a separate category, drawing on shared traits without fully merging.

SUBJECTIVE DEPRECIATIVE EVALUATION (SDE): Y is perceived as an inferior or flawed version of X, subject to the speaker's negative evaluation. It captures a speaker's critical stance, attributing substandard or undesirable qualities to Y in comparison to the prototypical X.

VAGUENESS (V): This value arises when Y is an entity whose nature is uncertain and has a tenuous relationship to X. It may be seen as a borderline or ambiguous member of the category X, suggesting fuzzy category distinctions.

3.4.1. *Kvazi(-)*

The analyzed sample of *kvazi(-)* formations reveals that this prefixoid is predominantly used in Croatian to negatively evaluate the referent, suggesting it fails to represent its category appropriately. This evaluative function, termed SUBJECTIVE DEPRECIATIVE EVALUATION, reflects instances where *kvazi(-)* highlights perceived deficiencies in the subject, as demonstrated in the example (8), where athletes are disparaged, possibly due to their lifestyle choices. Examples provided in this subsection, as well as in the following ones, derive from the aforementioned CLASSLA corpus.

- (8) *Najbolji naš igrač prošle sezone, igra većim srcem nego bilo tko drugi, [...] ne opija se po vikendima za razliku od 95% hrvatskih **kvazisportaša**.*

‘Our best player last season, he plays with a bigger heart than anyone else, [...] he doesn’t get drunk on weekends, unlike 95% of Croatian pseudo-athletes (lit. KVAZIathletes).’

Furthermore, a prototypical function of *kvazi(-)* is noted, viz. degree modifier-like behavior, in which *kvazi(-)* indicates the sense of ‘coming close’ to the concept expressed by the head noun. In such cases, *kvazi(-)* can be paraphrased with adverbs like *gotovo* ‘almost’ and *skoro* ‘nearly’, exemplifying a genuine approximation value, as seen in (9). We call this value “CLOSE-TO” EVALUATION.

- (9) *Europski sindikalni pokret zabrinut je zbog prijedloga o ekonomskom upravljanju [...] koji bi zemlje članice sveo na status **kvazikolonije**.*

‘The European trade union movement is concerned about the proposals on economic governance [...] which would reduce member countries to the status of a quasi-colony.’

Moreover, *kvazi(-)* occasionally conveys the value of FAKENESS, where it implies intentional deception, as in the case of fake sales tactics described in (10). Here, *kvazi(-)* indicates that X pretends to be genuine to mislead:

- (10) *[...] stvarno ne vidim smisla ovdje dodatno reklamirati **kvazi akcije** trgovina koje prvo dignu cijene 50% pa spuštaju 30%.*

‘I really don’t see the point of additionally advertising fake-sales (lit. KVAZIsales) of stores that first increase the prices by 50% and then lower them by 30%.’

Lastly, *kvazi(-)* is also found in expressions classified as IMITATION, where *kvazi-X* denotes an imitation or reproduction of X without the intention of being perceived as authentic, as illustrated by *kvazibolonjez* ‘pseudo-bolognese’ in (11), a plant-based version of the original Italian dish.

- (11) *I htjedoh reći – onaj **kvazibolonjez** od crvene leće je odličan, pogotovo kad se ohladi.*

‘And I wanted to say – that red lentil pseudo-bolognese (lit. KVAZIBolognese) is great, especially when it’s cold.’

Table 8 presents the semantic classification of the analyzed *kvazi(-)* formations.

Tab. 8: Semantic classification of the 500-token *kvazi(-)* sample

PREFIXOID	VALUE	TOKENS
<i>kvazi(-)</i>	SUBJECTIVE DEPRECIATIVE EVALUATION	492
	“CLOSE-TO” EVALUATION	4
	FAKENESS	2
	IMITATION	2

3.4.2. *Nadri(-)*

The analysis of 500 tokens containing *nadri(-)* reveals that 446 tokens (89.20% of the total) correspond to four specific nouns: *nadriliječnik* ('quack'), *nadripisar* ('quack scribe'), and their related nouns denoting the activities performed by such individuals, *nadriliječništvo* ('quackery') and *nadripisarstvo* ('scribal quackery'). In these contexts, *nadri(-)* refers to individuals who deceitfully engage in a profession for which they lack the requisite education and competence, or to the illicit activities carried out by such individuals, as illustrated in (12):

- (12) *U Statutu Hrvatske liječničke komore [...] stoji da Komora "promiče znanstvene postupke dijagnostike i liječenja a suzbija **nadriliječništvo**" (čl. 33. t. 7).*

'The Statute of the Croatian Medical Chamber [...] states that the Chamber "promotes scientific diagnostic and treatment procedures and suppresses quackery (lit. NADRI_{treatment})" (art. 33, cl. 7).'

As discussed earlier, in these predominantly legal contexts, *nadri(-)* conveys a specific, well-defined sense, which we term ILLEGALITY. Out of the 446 tokens of the aforementioned four nouns, 423 were annotated as ILLEGALITY due to their direct reference to the legal terminology. Although ILLEGALITY might overlap with FAKENESS or IMITATION (as a quack could be perceived as a fake doctor), there are nuanced differences. The criterion of intentionality, essential for FAKENESS (Masini & Micheli 2020), is not always clear-cut. For example, a quack doctor may deliberately pose as a real doctor for profit, as seen in (13), where Dulcamara from Donizetti's famous opera buffa *L'elisir d'amore* exploits the villagers' gullibility by selling wine as a supposed magical love potion:

- (13) *Ljubavni napitak slavna je komična opera, u kojoj je sjajno iznio komične karaktere likova, posebno **nadriliječnika** Dulcamaru, koji iskorištava seosku lakovjernost i dobrohotnost.*

'*L'elisir d'amore* is a famous comic opera, in which he brilliantly brought out the comic characters of the characters, especially the quack (lit. NADRI_{doctor}) Dulcamara, who exploits the gullibility and benevolence of the village.'

In this instance, it might be possible to classify the occurrence as an instance of FAKENESS, given the assumed intentionality of deception. However, in other instances, such as in (14), there is no clue on whether subjects in question are intentionally posing as licensed

professionals, or they are categorized as quacks by contemporary standards while in their era they were recognized in fact as “medical” figures.

- (14) *U srednjem vijeku dolazi do stagnacije u području liječenja zubi jer se tim poslom počinju baviti brijači i drugi **nadriliječnici**.*

‘In the Middle Ages, there was stagnation in the field of dental treatment, as barbers and other quacks (lit. NADRIDoctors) began to deal with this business.’

To avoid ambiguous interpretations and given the strong association of *nadri* with legal contexts, ILLEGALITY is deemed the most accurate classification for these occurrences. Finally, in certain contexts, *nadri(-)* is used to negatively assess a subject, devoid of any reference to illegality or clues about the subject’s formal education (15) or even confirming that the subject, in fact, holds a relevant degree (16). In these cases, *nadri(-)* reflects SUBJECTIVE DEPRECIATIVE EVALUATION.

- (15) *Nije ovo nikakva kritika tebi već više debilnom jeziku i još debilnijim našim **nadri jezikoslovcima**.*

‘This is not a criticism of you, but of a stupid language and our even more stupid quack linguists (lit. NADRILinguists).’

In (15), the linguists in question may have formal degrees, yet their ideas are considered irrelevant or nonsensical. Conversely, in (16), it is explicitly stated that “quacks” possess a degree, but, for instance, their treatment methods are viewed negatively.

- (16) [...] *najviše **nadriliječnika** se skriva iza diplome liječnika.*

‘[...] most quacks (lit. NADRIDoctors) are hiding behind a doctor’s degree.’

As a result, the 23 tokens from the group of 446 (*nadriliječnik*, *nadriliječništvo*, *nadripisar*, *nadripisarstvo*) and the remaining 54 tokens were annotated as SUBJECTIVE DEPRECIATIVE EVALUATION.

Table 9 presents the semantic classification of the 500 analyzed *nadri(-)* expressions.

Tab. 9: Semantic classification of the 500-token *nadri(-)* sample

PREFIXOID	VALUE	TOKENS
<i>nadri(-)</i>	ILLEGALITY	423
	SUBJECTIVE DEPRECIATIVE EVALUATION	77

3.4.3. *Nazovi(-)*

Nazovi(-), just like *nadri(-)*, makes part of a legal terminology and formation such as *nazoviiliječništvo* ‘quackery’ is synonymous to *nadriliječništvo* ‘quackery’, as it indicates the illegal activity of providing medical assistance by a non-professional person. However, in the analyzed sample, no tokens of *nazoviliječništvo* ‘quackery’ or *nazoviliječnik* ‘quack’ were identified, and only a single occurrence of *nazoviliječništvo* was found in the entire corpus, indicating the rarity of such formation. This scarcity could be attributed to a phenomenon known as *statistical preemption* (Boyd & Goldberg 2011), where the frequent occurrence of *nadriliječnik* in contexts in which also *nazoviliječnik* would have been adequate has led to its entrenchment (due to its frequency), thereby preempting the use of *nazoviliječnik*. Regarding *nazovi(-)*, as illustrated in (17), we observe how the prefix in most cases carries a value of SUBJECTIVE DEPRECIATIVE EVALUATION.

- (17) *Granica dobrog ukusa debelo je prekoračena u ovom **nazovi filmu**.*

‘The limit of good taste is grossly overstepped in this pseudo-film (lit. NAZOVI film).’

Additionally, *nazovi(-)*, due to its origin as an imperative form, can express a function no other analyzed prefix carries, namely one of a hedge (Lakoff 1973). More precisely, *nazovi(-)* can be used as a (non-morphological) mechanism of coming close or matching with what is intended to be signified, as seen in (18) and (19).

- (18) *Današnjim klicima rat je pomrsio račune, [...] i nitko ih ne bi trebao kriviti za tu **nazovi hladnoću**.*

‘For today’s kids, the war messed up their plans [...], and nobody should blame them for that so-called coldness (lit. NAZOVI coldness).’

- (19) *Išla sam na Hitnu, dobila injekciju, kroz neku **nazovi bocu** sam disala. Dr. mi je rekla da sam pobrala neku bakteriju.*

‘I went to the emergency room, got an injection, I breathed through some type of a bottle (lit. NAZOVI bottle). Doctor told me that I’ve cached some bacteria.’

In these cases, we argue that *nazovi(-)* does not serve carry subjective depreciative value towards the concept indicated by the noun but serves as a hedge, collocating the designated noun as an atypical member of its category or even questioning its very categorization as the head noun. It is usually referred to such a value as VAGUENESS.

Finally, instances of FAKENESS were identified, as in (20), in which it is clear that the modified noun indicates individuals who are intentionally trying to imitate being a patriot without actually being a genuine patriot.

- (20) *Mene je beskrajno stid zbog njih, ali sam istovremeno ponosan sto ne pripadam toj bestidnoj sorti **nazovi rodoljuba**.*

‘I am beyond of ashamed because of them, but at the same time I am proud that I do not belong to that shameless breed of so-called patriots (lit. NAZOVI patriots).’

Lastly, one example of KIN-CATEGORIZATION was found. The formation *nazovi brak* ‘pseudo-marriage’, illustrated in (21), makes a reference to civil unions between same-sex partners, which, in this context, carry a clearly pejorative connotation.

- (21) *Sve druge kombinacije što ih nameće moderni svijet [...] neprirodne su prema biblijskim načelima: poligamija, **nazovibrakovi** između pripadnika/ca istog spola.*

‘All other combinations imposed by the modern world [...] are unnatural according to biblical principles: polygamy, pseudo-marriages (lit. NAZOVI marriages) between same-sex partners.’

Table 10 presents the semantic classification of the 500 analyzed *nazovi(-)* tokens.

Tab. 10: Semantic classification of the 500-token *nazovi(-)* sample.

PREFIXOID	VALUE	TOKENS
<i>nazovi(-)</i>	SUBJECTIVE DEPRECIATIVE EVALUATION	334
	VAGUENESS	92
	FAKENESS	73
	KIN-CATEGORIZATION	1

3.4.4. *Pseudo(-)*

Among the four examined prefixoids, *pseudo(-)* exhibits the highest frequency in scientific and technical language. In these contexts, *pseudo(-)* functions as a left constituent of neoclassical compounds, usually combining with neoclassical Final Combining Forms as *pseudonim* ‘pseudonym’ or a non-classical scientific terms such as *pseudohipokalcemija* ‘pseudohypocalcemia’. Here, *pseudo(-)* conveys a classifying (privative) sense, signaling that the modified noun does not belong to the category of its head noun. Given that these terms are predominantly part of scientific jargon, they do not convey a subjective negative

evaluation of the subject, but rather indicate an exclusion from the category, as illustrated in (22).

- (22) *Transabdominalnom ultrazvučnom pretragom se postavlja sigurna dijagnoza **pseudo-gravidnosti** na temelju odsutnosti placentoma, fetusa i fetalnih membrana.*

‘A transabdominal ultrasound examination establishes a safe diagnosis of pseudo-pregnancy based on the absence of the placentome, fetus and fetal membranes.’

These instances are annotated as conveying SCIENTIFIC NON-EVALUATIVE (SNE) meaning, which lies outside the scope of evaluative morphology since no explicit evaluation is expressed. Another identified value is SUBJECTIVE DEPRECIATIVE EVALUATION, where *pseudo(-)* does not indicate the exclusion from the head’s category (thus, *pseudoX* is, in fact, X) but the head’s qualitative depreciation, serving as a derogatory marker, as shown in (23). This value often appears in formations with native elements rather than scientific terms, as in *pseudo poezija* (‘pseudo poetry’), which refers to poorly written poetry devoid of artistic merit.

- (23) *Loša erotska **pseudo poezija** na stranu, budite liberalni sa isprobavanjem novih stvari u postelji.*

‘Bad erotic pseudo poetry aside, be liberal with trying new things in bed.’

Finally, an instance has been identified where it is not possible (at least with certainty) to determine whether the occurrence exemplifies category-inclusion or category-exclusion. In this context, as highlighted by Vassiliadou et al. (2023), “semantic vagueness meets [...] subjective vagueness: as *pseudo(-)* exploits the existence of borderline cases and underlines the negative side of X, it questions the very categorization of X”. In the example provided in (24), it is unclear whether *pseudoljubav* ‘pseudo-love’ denotes a form of love that is not genuine (an imitation), a depreciated form of love, or even calls into question the concept of love itself.

- (24) *Nedostatak ljubavi, koji se često očituje i kroz uvjetovanost ljubavi i razne oblike **pseudoljubavi**, može pridonijeti razvoju otuđenosti [...]*

‘Lack of love, which is often manifested through the conditionality of love and various forms of pseudo-love, can contribute to the development of alienation [...]

This example, annotated as VAGUENESS, highlights the complex interplay between semantic and subjective vagueness, underscoring how *pseudo(-)* can destabilize clear-cut

categorizations, particularly in contexts where the boundaries of concepts are fluid or disputed.

Table 11 presents the semantic classification of the 500 analyzed *pseudo(-)* expressions.

Tab. 11: Semantic classification of the 500-token *pseudo(-)* sample.

PREFIXOID	VALUE	TOKENS
<i>pseudo(-)</i>	SCIENTIFIC NON-EVALUATIVE	198
	SUBJECTIVE DEPRECIATIVE EVALUATION	301
	VAGUENESS	1

3.4.5. Distribution of Semantic Values

Following an analysis of the distribution of semantic values expressed by each of the four prefixoids within the 500-token sample per prefixoid, it is pertinent to visualize these distributions across all prefixoids. Table 12 illustrates the distribution of semantic values among the derivatives constructed with the four prefixoids.

Tab. 12: Distribution of semantic values among derivatives (500 tokens per prefixoid). For each prefixoid and semantic value, the token count is provided, with the type count indicated in parentheses.

VALUE PREFIXOID	CTE	F	ILL	IM	K-C	SDE	V	SNE
<i>kvazi(-)</i>	4 (3)	2 (1)	/	2 (2)	/	492 (52)	/	/
<i>nadri(-)</i>	/	/	423 (6)	/	/	77 (10)	/	/
<i>nazovi(-)</i>	/	73 (12)	/	/	1 (1)	334 (41)	92 (9)	/
<i>pseudo(-)</i>	/	/	/	/	/	301 (19)	1 (1)	198 (36)

Observing the table, it can be confirmed how all four prefixoids are polyfunctional, with the distribution of the derivatives in relation to semantic value varying significantly among the prefixoids. Prefixoids *kvazi(-)* and *nazovi(-)* convey the highest number of semantic values (4), while *pseudo(-)* conveys 3, and *nadri(-)* only 2 values. To quantify the prefixoid polyfunctionality, we consider both the proportion of semantic values shared between them and the frequency of lexical realization (number of types) of these values, as per Huyghe et al. (2023). In alignment with the methodology of Salvadori, Varvara & Huyghe

(2024), we calculate the Hill-Shannon diversity index D^{14} (cf. Roswell, Dushoff & Winfree 2021) utilizing the *MeanRarity*¹⁵ package. According to D , *nazovi(-)* ($D = 2.56$) displays the highest diversity, followed by *pseudo(-)* ($D=2.06$) and *nadri(-)* ($D=1.94$), while *kvazi(-)* ($D=1.55$) is determined to be the least diverse.

While the necessity for a scalar evaluation of rivalry, that is, a metric¹⁶ to delineate the degrees of rivalry for accurately assessing the intensity of a rivalry relationship, is apparent in occurrences of rivalry among multiple affixes (Huyghe et al. 2023; Salvadori, Varvara & Huyghe 2024), the present study delivers a comparatively clear scenario. It can be noted how five semantic values are associated exclusively with one specific prefixoid: “CLOSE-TO” EVALUATION and IMITATION to *kvazi(-)*, ILLEGALITY to *nadri(-)*, KIN-CATEGORIZATION to *nazovi(-)*, and SCIENTIFIC NON-EVALUATIVE to *pseudo(-)*. Conversely, the values of FAKENESS are expressed by *kvazi(-)* and *nazovi(-)*, VAGUENESS is expressed by *nazovi(-)* and *pseudo(-)*¹⁷, and, lastly, SUBJECTIVE DEPRECIATIVE EVALUATION is conveyed by all four prefixoids. Consequently, based on the examination of a 500-token sample for each prefixoid, it can be inferred that competition among the four prefixoids predominantly emerges within contexts characterized by the speaker’s negative evaluation of a subject. In these instances, prefixoids are utilized to signify undesirable attributes in contrast to the prototypical instance. Conversely, for other, less frequent semantic values, distinct preferences concerning the selection of the prefixoid become evident. This observation does not come as a surprise, as the value referred to as SUBJECTIVE DEPRECIATIVE EVALUATION is, in fact, a very frequent part of language use, whereas functions like KIN-CATEGORIZATION or VAGUENESS appear less frequently in common discourse. The increased prevalence of

¹⁴ Unlike the conventional Hill index, Hill-Shannon diversity emphasizes neither rare nor common species (in this context, semantic values). It is defined with the exponent that determines the rarity scale on which the mean is taken (l) of 0 and calculated with the base of the natural logarithm, e , raised to the power of the traditional Shannon entropy index (for detailed formalizations, see Roswell, Dushoff & Winfree 2021).

¹⁵ https://mikeroswell.github.io/MeanRarity/articles/Using_MeanRarity.html (accessed 12 November 2024).

¹⁶ Several methods that could (either implicitly or explicitly) account for affix rivalry have been explored. Fernández-Domínguez (2017) introduces a competition index that evaluates the prevalence of a derivative in relation to its rivals. Guzmán Naranjo & Bonami (2023) measure the similarity between different word-formation representations by calculating average vector offsets between the distributional representations of derivatives and their bases. Lastly, Salvadori, Varvara, & Huyghe (2024) utilize semantic annotation of derivatives to estimate the number of (un)shared functions among rival affixes, applying incidence-based measures (Sørensen index) and abundance-based measures (Percentage similarity coefficient).

¹⁷ However, it has to be noted that only one instantiation of VAGUENESS with *pseudo(-)* is observed so it could be argued that VAGUENESS is, in fact, exclusively a property of *nazovi(-)*.

negative evaluative contexts, consequently, provides more opportunities for competitive usage of these prefixoids.

4. Conclusions

This usage-based study has introduced and applied several statistical methods to analyze four competing <PREF_{APPRX} + noun> constructions, revealing that the combinations of the analyzed prefixoids and nouns are not entirely unconstrained.

The analysis reaffirms that affix rivalry is a gradient phenomenon, necessitating the examination of a comprehensive range of factors to elucidate its nature. First, it was shown that prefixoids vary significantly in their productivity, with *nazovi(-)* identified as the most productive and *nadri(-)* the least productive, exhibiting a productivity difference of 3.72 times between the two (based on Potential Productivity). Second, the collocational preferences of the prefixoids were scrutinized. The hypothesis of a significant semantic similarity between *kvazi(-)* and *nazovi(-)*, premised on a great number of shared bases between the two, was corroborated by MDCA-based Correspondence analysis. This analysis revealed that the two prefixoids cluster together, with ellipses embracing the collexemes that are associated with two prefixoids overlapping almost perfectly, thereby suggesting a functional and semantic similarity between them. In contrast, distinct semantic behavior of *nadri(-)* and *pseudo(-)* was observed, with no overlap in their collocational preferences. The distinct clustering of *pseudo(-)* with scientific and medical terms corroborates the prefixoid's specialized role in denoting false or imitative phenomena within technical discourse, while the separation of *nadri(-)* implies another specialized usage, viz. denoting fraudulent or unqualified individuals in professional roles. Lastly, a sample of 500 tokens for each prefixoid was analyzed, yielding several semantic readings and proposing a classification based on the typical values conveyed by these prefixoids.

Particularly intriguing are the findings regarding *kvazi(-)* and its possibility to express fakeness. As already noted, the reference literature examines *quasi(-)* in English (almost) exclusively as an approximative prefixoid. Bauer, Lieber & Plag (2013) highlight how *quasi(-)*, unlike *pseudo(-)*, lacks the element of falseness, while Dixon (2014: 171) uses a

contrastive pair to illustrate the falseness *pseudo(-)* lacks to convey: whereas *quasi-cripple* represents “someone who has some small thing wrong with them (say, missing two fingers from one hand) but not really so serious to justify the label ‘cripple’”, a *pseudo-cripple* indicates “someone who has nothing at all wrong with them but pretends to be a cripple (perhaps, so that they can take part in the para-Olympics)”. Cappelle, Daus & Hartmann (2023) build upon this by asserting that *quasi(-)* predominantly has an approximative meaning, with *pseudo(-)* being characterized as disproximative. In the end, however, they acknowledge that *quasi(-)* can also convey disproximation, but the impression remains that that reading according to the authors is rather infrequent. In fact, Masini, Norde & Van Goethem (2023: 11) in their introduction to the Special Issue stress how according to Cappelle, Daus & Hartmann (2023) the “idea of deficiency (...) is not prominent in *quasi-*”. In this study, we assert that *kvazi(-)* in Croatian can convey both what Cappelle, Daus & Hartmann (2023) define as *disproximation* (understood as a more inclusive, wider term), and *fakeness/falseness* (understood as “intentional act of deception or obfuscation” (Bauer, Lieber & Plag 2013: 416)). On a general note regarding semantic annotation, once again *kvazi(-)* and *nazovi(-)* display the most overlap, particularly in their shared use of SUBJECTIVE DEPRECIATIVE EVALUATION. Both prefixoids often function as tools of criticism, emphasizing the insufficiency or pretentiousness of the subject. There is also a notable, albeit less frequent, overlap between *kvazi(-)* and *pseudo(-)* in expressing IMITATION, highlighting their shared role in marking resemblance without authenticity. *Nadri(-)*, while sharing the FAKENESS value with the other prefixoids, is more narrowly focused on legal contexts, limiting its broader evaluative overlap. *Pseudo(-)*, with its scientific leanings, remains distinct but conceptually adjacent to *kvazi(-)* in contexts where imitation is highlighted without necessarily invoking deceit. Overall, the prefixoids display a complex web of relationships, yet each prefixoid also carves out a specific niche, balancing between shared semantic functions and unique, context-driven applications.

The adopted approach is expected to offer novel insights into the study of approximative prefixoids in Croatian. However, needless to say, the presented findings shed some light only on selected aspects of this previously overlooked phenomenon, and further research is needed to enhance our understanding of the analyzed prefixoids. An intriguing question remains regarding the original nature of these constructions. For instance, in the case

nadri(-), while it cannot be confirmed due to limited resources, one could hypothesize that *nadriliječnik* ‘quack’ and the related term *nadriliječništvo* ‘quackery’ exemplify what Rainer (2013) refers to as a *leader word*, i.e. a word which serves as a model for the formation of new words and the development of a new construction. If this is indeed the case, it would suggest that the construction has witnessed a broadening of meaning (from what we defined as ILLEGALITY to SUBJECTIVE DEPRECIATIVE EVALUATION), giving rise to a more general approximative construction. However, it is equally, if not more, plausible that the reverse has occurred: a construction initially conveying SUBJECTIVE DEPRECIATIVE EVALUATION has become more specialized, entering legal terminology. For now, this question remains open. Furthermore, from a methodological perspective, a comprehensive understanding of rivalry resolution necessitates a holistic view of the potential differences between affixes (Huyghe & Varvara 2023). Future research should therefore incorporate regression modeling or classifier approaches, using the factors presented here along with additional variables. Moreover, due to prefixoids’ different etymologies and presumably diverging periods in which they enter the lexicon, a diachronic study, comparing results from different time periods, may be relevant to account for this case of affix rivalry. However, given the scarcity of resources for Croatian (sufficiently large diachronic corpus or a tool such as Google Books Ngram Viewer), this task poses significant challenges. In addition, prefixoids’ preferences regarding the etymology of their bases should be investigated. While some authors suggest that *kvazi(-)* is found with non-native words while *nadri(-)* and *nazovi(-)* combine with Croatian words (Ham 2015), a corpus-based study is needed to verify these claims. Finally, similar to Vassiliadou et al. (2023), instances of prefixoids appearing in quotation marks or parentheses, as seen for all four prefixoids, should be explored to determine whether they reflect merely an orthographic variation or actual uncertainty regarding the modified head’s status (e.g., whether a term like *(nadri)liječnik* denotes a legitimate doctor or a potential quack).

References

- Amiot, Dany & Dejan Stosic. 2022. Evaluative morphology: From evaluation to approximation and semi-categorization. In H       Vassiliadou & Marie Lammert (eds.), *A crosslinguistic perspective on clear and approximate categorization*, 53–94. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Aronoff, Mark. 2019. Competitors and alternants in linguistic morphology. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Lusch  tzky (eds.), *Competition in inflection and word-formation*, 39–66. Berlin: Springer.
- Aronoff, Mark. 2023. Three ways of looking at morphological rivalry. *Word Structure* 16(1). 49–62.
- Baayen, Harald & Rochelle Lieber. 1991. Productivity and English derivation: A corpus-based study. *Linguistics* 29. 801–844.
- Baayen, Harald. 1992. A quantitative approach to morphological productivity. In Geert Booij & Jaap van Marle (eds.), *Yearbook of Morphology 1991*, 109–149. Dordrecht: Kluwer.
- Baayen, Harald. 2001. *Word frequency distributions*. Dordrecht: Kluwer.
- Baayen, Harald. 2009. Corpus linguistics in morphology: morphological productivity. In Anke L         & Merja Kyt     (eds.), *Corpus linguistics: An international handbook*, 900–919. Berlin & New York: De Gruyter.
- Ba      , Marta, Miquel Espl        , Mikel L. Forcada, Cristian Garc          , Taja Kuzman, Nikola Ljube      , Rik van Noord, Leopoldo Pla Sempere, Gema Ram            , Peter Rupnik, V     Suchomel, Antonio Toral, Tobias van der Werff & Jaume Zaragoza. 2023. *Croatian web corpus MaCoCu-hr 1.0*. Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1516> (accessed 27 August 2024).
- Bar      , J        , Renata Enghels, Quentin Feltgen, Sven Van Hulle & Peter Lauwers. in prep. Productivity in diachrony. In Adam Ledgeway, Edith Aldridge, Anne Breitbarth, Katalin E. Kiss, Joseph Salmons & Alexandra Simonenko (eds.), *The Wiley Blackwell Companion to Diachronic Linguistics*. Oxford: Wiley-Blackwell.
- Bari    , Eugenija. 1979. Dosada               slo         u hrvatskom i srpskom jeziku [Previous study of compounds in the Croatian and Serbian language]. *Rasprave* 4–5. 17–29.
- Bari    , Eugenija. 1980. Imeni       slo         s glagolskim prvim dijelom [Noun compounds with the verbal first part]. *Rasprave* 6–7. 17–30.
- Baroni, Marco & Stephan Evert. 2014. *The zipfr package for lexical statistics: A tutorial introduction*. <https://zipfr.r-forge.r-project.org/materials/zipfr-tutorial.pdf> (accessed 17 July 2024).
- Batini    , Mia, Marijana Kres       & Anita Pavi            . 2015. The intensifying function of modal particles and modal elements in a cross–linguistic perspective. *Rasprave* 41(1). 1–27.
- Bauer, Laurie, Rochelle Lieber & Ingo Plag. 2013. *The Oxford reference guide to English morphology*. Oxford: Oxford University Press.
- Bauer, Laurie, Salvador Valera & Ana D              . 2010. Affixation vs. conversion: The resolution of conflicting patterns. In Franz Rainer, Wolfgang U. Dressler, Dieter Kastovsky & Hans Christian Lusch  tzky (eds.), *Variation and change in morphology*, 15–32. Amsterdam: Benjamins.
- Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.
- Benz      , Jean-Paul. 1973. *L’analyse des donn        . 2. L’analyse des correspondances*. Paris: Bordas.
- Boyd, Jeremy K. & Adele E. Goldberg. 2011. Learning what *not* to say: the role of statistical preemption and categorization in *a*-adjective production. *Language* 87(1). 55–83.

- Bybee, Joan L. 1985. *Morphology: A study of the relation between meaning and form*. Amsterdam: Benjamins.
- Cappelle, Bert, Pascal Denis & Mikaela Keller. 2018. Facing the facts of fake: A distributional semantics and corpus annotation approach. In Beate Hampe & Susanne Flach (eds.), *Yearbook of the German Cognitive Linguistics Association* 6, 9–42. Berlin & New York: De Gruyter.
- Cappelle, Bert, Robert Daus & Stefan Hartmann. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and ‘disproximation’. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 52–75.
- Corbin, Danielle. 1987. *Morphologie dérivationnelle et structuration du lexique*. Tübingen: Max Niemeyer.
- Core Team. 2021. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org/> (accessed 10 June 2024).
- Covington, Michael A. & Joe D. McFall. 2010. Cutting the Gordian knot: The Moving-Average Type–Token Ratio (MATTR). *Journal of Quantitative Linguistics* 17(2). 94–100.
- Cúneo, Paola. 2015. Toba. In Nicola Grandi & Livia Körtevelyessy (eds.), *The Edinburgh handbook of evaluative morphology*, 625–633. Edinburgh: Edinburgh University Press.
- Desagulier, Guillaume. 2014. Visualizing distances in a set of near-synonyms: *rather*, *quite*, *fairly*, and *pretty*. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus methods for semantics: Quantitative studies in polysemy and synonymy*, 145–178. Amsterdam: Benjamins.
- Desagulier, Guillaume. 2015. Forms and meanings of intensification: a multifactorial comparison of *quite* and *rather*. *Anglophonia – French Journal of English Linguistics* 20. <https://journals.openedition.org/anglophonia/558> (accessed 8 July 2024).
- Dixon, Robert M. W. 2014. *Making new words: Morphological derivation in English*. Oxford: Oxford University Press.
- Dressler, Wolfgang U. & Lavinia M. Barbaresi. 1994. *Morphopragmatics. Diminutives and intensifiers in Italian, German, and other languages*. Berlin & New York: De Gruyter.
- Ellis, Nick C. & Fernando Ferreira-Junior. 2009. Constructions and their acquisition: Island and the distinctiveness of their occupancy. Francisco José Ruiz de Mendoza Ibáñez (ed.), *Annual review of cognitive linguistics: Volume 7*, 187–220. Amsterdam: Benjamins.
- Ellis, Nick C. 2007. Language acquisition as rational contingency learning. *Applied Linguistics* 27(1). 1–24.
- Evert, Stephan & Anke Lüdeling. 2001. Measuring morphological productivity: Is automatic preprocessing sufficient? In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie & Shereen Khoja (eds.), *Proceedings of the Corpus Linguistics 2001 conference*, 167–175. Lancaster: University Centre for Computer Research on Language.
- Evert, Stephan & Marco Baroni. 2022. Package ‘zipfR’. *Statistical Models for Word Frequency Distributions*. <https://cran.r-project.org/web/packages/zipfR/zipfR.pdf> (accessed 9 July 2024).
- Fernández-Domínguez, Jesús. 2013. Morphological productivity measurement: Exploring qualitative versus quantitative approaches. *English Studies* 94(4). 422–447.
- Fernández-Domínguez, Jesús. 2017. Methodological and procedural issues in the quantification of morphological competition. In Juan Santana-Lario & Salvador Valera (eds.), *Competing Patterns in English Affixation*, 67–117. Lausanne: Peter Lang.
- Flach, Susanne. 2021. *Collostructions: An R implementation for the family of collostructional methods* (Version 0.2.0) [R Script].

- Gaeta, Livio & Davide Ricca. 2006. Productivity in Italian word formation: A variable-corpus approach. *Linguistics* 44(1). 57–89.
- Gaeta, Livio & Davide Ricca. 2015. Productivity. In Peter O. Müller, Ingeborg Ohnheiser, Susan Olsen & Franz Rainer (eds.), *Word-formation: An international handbook of the languages of Europe*, 842–858. Berlin & New York: De Gruyter.
- Gardani, Francesco, Franz Rainer & Hans Christian Luschützky. 2019. Competition in morphology: A historical outline. In Franz Rainer, Francesco Gardani, Wolfgang U. Dressler & Hans Christian Luschützky (eds.), *Competition in inflection and word-formation*, 3–36. Berlin: Springer.
- Goldberg, Adele E., Devin M. Casenhiser & Nitya Sethuraman. 2004. Learning argument structure generalizations. *Cognitive Linguistics* 15(3). 289–316.
- Grandi, Nicola. 2017. Evaluatives in Morphology. In *Oxford Research Encyclopedia of Linguistics* (accessed 9 August 2024)
- Grandi, Nicola & Livia Körvélyessy. 2015. Introduction: why evaluative morphology. In Nicola Grandi & Livia Körvélyessy (eds.), *Edinburgh handbook of evaluative morphology*, 3–20. Edinburgh: Edinburgh University Press.
- Grandi, Nicola. 2002. *Morfologie in contatto: le costruzioni valutative nelle lingue del Mediterraneo*. Milan: FrancoAngeli.
- Greenacre, Michael. 2017. *Correspondence analysis in practice. Third edition*. Boca Raton: Chapman & Hall/CRC Press.
- Gries, Stefan Th. & Anatol Stefanowitsch. 2004. Extending collostructional analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics* 9(1). 97–129.
- Gries, Stefan Th. 2015. More (old and new) misunderstandings of collostructional analysis: On Schmid and Küchenhoff (2013). *Cognitive Linguistics* 26(3). 505–536.
- Gries, Stefan Th. 2019. *Ten lectures on corpus-linguistic approaches: Applications for usage-based and psycholinguistic research*. Leiden: Brill.
- Gries, Stefan Th. 2022. Toward more careful corpus statistics: uncertainty estimates for frequencies, dispersions, association measures, and more. *Research Methods in Applied Linguistics* 1(1).
- Gries, Stefan Th. 2023. Overhauling collostructional analysis: Towards more descriptive simplicity and more explanatory adequacy. *Cognitive Semantics* 9(3). 351–386.
- Guzmán Naranjo, Matías & Bonami, Olivier. 2023. A distributional assessment of rivalry in word formation. *Word Structure* 16(1). 87–114.
- Ham, Sanda. 2015. Makro, mikro [Macro, micro]. *Jezik: časopis za kulturu hrvatskoga književnog jezika* 62. 144–145.
- Hartmann, Stefan. 2018. Derivational morphology in flux: A case study of word-formation change in German. *Cognitive Linguistics* 29(1). 77–119.
- Hein, Katrin & Annelen Brunner. 2020. Why do some lexemes combine more frequently than others? An empirical approach to productivity in German compound formation. In Jenny Audring, Nikos Koutsoukos & Christina Manouilidou (eds.), *Rules, patterns, schemas and analogy. Proceedings of Mediterranean Morphology Meetings 12*, 28–41. Patras: University of Patras.
- Huyghe, Richard, Alizée Lombard, Justine Salvadori & Sandra Schwab. 2023. Semantic rivalry between French deverbal neologisms in *-age*, *-ion* and *-ment*. In Sven Kotowski & Ingo Plag (eds.), *The Semantics of Derivational Morphology. Theory, Methods, Evidence*, 125–158. Berlin: De Gruyter.

- Huyghe, Richard & Rossella Varvara. 2023. Affix rivalry: Theoretical and methodological challenges. *Word Structure* 16(1). 1–23.
- Ivšić, Stjepan. 1906–1907. Nešto o riječima složenima s *nadri*- [Something about words composed with *nadri*-]. *Nastavni vjesnik* 15. 525–527.
- Jojić, Ljiljana (ed.). 2015. *VRH – Veliki rječnik hrvatskoga standardnog jezika* [VRH – Large Dictionary of Croatian Standard Language]. Zagreb: Školska knjiga.
- Klajn, Ivan. 2002. *Tvorba reči u savremenom srpskom jeziku. Prvi deo: slaganje i prefiksacija* [Word formation in the contemporary Serbian language. Part first: composition and prefixation]. Beograd – Novi Sad: Zavod za udžbenike i nastavna sredstva – Institut za srpski jezik SANU – Matica srpska.
- Körtvélyessy, Livia & Pavol Štekauer (eds.). 2011. Diminutives and augmentatives in the languages of the world. *Lexis: e-journal in English lexicology* 6. 5–25.
- Körtvélyessy, Livia. 2015. *Evaluative morphology from a cross-linguistic perspective*. Cambridge: Cambridge Scholars Publishing.
- Kuna, Branko. 2006. Nazivlje u tvorbi riječi [Terminology in Word-Formation]. *Filologija* 46. 165–182.
- Kuzman, Taja, Igor Mozetič & Nikola Ljubešić. 2023. Automatic genre identification for robust enrichment of massive text collections: Investigation of classification methods in the era of large language models. *Machine Learning and Knowledge Extraction* 5(3). 1149–1175.
- Lacić, Ivan. 2024a. A corpus-based study of maximizer–adjective patterns in Croatian. *Language Sciences* 102.
- Lacić, Ivan. 2024b. An insight into the Croatian degree modifier paradigm and its clustering profiles. *Suvremena lingvistika* 50(97). 85–112.
- Lakoff, George. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic* 2. 458–508.
- Levshina, Natalia. 2015. *How to do linguistics with R. Data exploration and statistical analysis*. Amsterdam: Benjamins.
- Liao, Shengyu, Stefan Th. Gries, & Stefanie Wulff. 2024. Transfer five ways: applications of multiple distinctive collxeme analysis to the dative alternation in Mandarin Chinese. *Corpus Linguistics and Linguistic Theory*, aop.
- Ljubešić, Nikola & Davor Lauc. 2021. BERTić - the transformer language model for Bosnian, Croatian, Montenegrin and Serbian. In Bogdan Babych (ed.), *Proceedings of the 8th Workshop on Balto-Slavic Natural Language Processing*, 37–42. Association for Computational Linguistics.
- Ljubešić, Nikola & Filip Klubička. 2014. {bs,hr,sr}WaC –Web corpora of Bosnian, Croatian and Serbian. In Felix Bildhauer & Roland Schäfer (eds.), *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*, 29–35. Association for Computational Linguistics.
- Ljubešić, Nikola & Peter Rupnik, Taja Kuzman. 2024. *Croatian web corpus CLASSLA-web.hr 1.0*. Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1929> (accessed 10 May 2024).
- Ljubešić, Nikola & Taja Kuzman. 2024. CLASSLA-web: Comparable Web Corpora of South Slavic Languages Enriched with Linguistic and Genre Annotation. *arXiv:2403.12721*.
- Masini, Francesca & M. Silvia Micheli. 2020. The morphological expression of approximation: The emerging *simil*- construction in Italian. *Word Structure* 13(3). 371–402.

- Masini, Francesca, Muriel Norde & Kristel Van Goethem. 2023. Approximation in morphology: A state of the art. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 1–26.
- Merlini Barbaresi, Lavinia. 2015. Evaluative morphology and pragmatics. In Nicola Grandi & Livia Körvélyessy (eds.), *Edinburgh handbook of evaluative morphology*, 32–42. Edinburgh: Edinburgh University Press.
- Nagano, Akiko, Alexandra Bagasheva & Vincent Renner. 2024. Towards a competition-based word-formation theory. Core research questions and major hypotheses. In Alexandra Bagasheva, Akiko Nagano & Vincent Renner (eds.), *Competition in word-formation*, 1–31. Amsterdam: Benjamins.
- Napoli, Maria & Miriam Ravetto. 2017. Exploring intensification: synchronic, diachronic and cross-linguistic perspectives. Amsterdam: Benjamins.
- Nigoević, Magdalena. 2020. *Intenzifikacija u jeziku: s primjerima iz hrvatskog i talijanskog jezika* [Intensification in language: With examples from the Croatian and Italian Language]. Split: Filozofski fakultet Sveučilišta u Splitu.
- Norde, Muriel. 2009. *Degrammaticalization*. Oxford: Oxford University Press.
- Plag, Ingo. 1999. *Morphological productivity: Structural constraints in English derivation*. Berlin & New York: De Gruyter.
- Proisl, Thomas. 2022. *Use words, not constructions!* A new perspective on the unit of analysis in collostructional analysis. *International Journal of Corpus Linguistics* 27(3). 349–379.
- Rainer, Franz. 2013. Formación de palabras y analogía. Aspectos diacrónicos. In Isabel Pujol Payet (ed.), *Formación de palabras y diacronía*, 141–172. A Coruña: Universidade da Coruña.
- Roswell, Michael, Jonathan Dushoff & Rachael Winfree. 2021. A conceptual guide to measuring species diversity. *Oikos* 130(3). 321–338.
- Salvadori, Justine, Rossella Varvara & Richard Huyghe (2024). Measuring affix rivalry as a gradient relationship. In Alexandra Bagasheva, Nagano, Akiko & Vincent Renner (eds.), *Competition in word-formation*, 104–138. Amsterdam: Benjamins.
- Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27(3). 379–423.
- Šonje, Jure (ed.). 2000. *Rječnik hrvatskoga jezika* [Dictionary of the Croatian language]. Zagreb: Leksikografski zavod Miroslav Krleža – Školska knjiga.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.
- Stefanowitsch, Anatol & Susanne Flach. 2020. ‘Too big to fail but big enough to pay for their mistakes’: A collostructional analysis of the patterns [too ADJ to V] and [ADJ enough to V]. In Gloria Pastor Corpas & Jean-Pierre Colson (eds.), *Computational phraseology*, 247–272. Amsterdam: Benjamins.
- Stefanowitsch, Anatol. 2013. Collostructional analysis. In Thomas Hoffman & Graeme Trousdale (eds.), *The Oxford Handbook of Construction Grammar*, 209–306. Oxford: Oxford University Press.
- Sundquist, John D. 2020. Productivity, richness, and diversity of light verb constructions in the history of American English. *Journal of Historical Linguistics* 10(3). 349–388.
- Tafra, Branka & Petra Košutar. 2009. Rječotvorni modeli u hrvatskom jeziku [Word-Formation models in the Croatian language]. *Suvremena lingvistika* 35(67). 87–107.

- Terčon, Luka & Nikola Ljubešić. 2023. *Word embeddings CLARIN.SI-embed.hr 2.0*. Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1790> (accessed 21 June 2024).
- Van Goethem, Kristel & Muriel Norde. 2020. Extravagant “fake” morphemes in Dutch. Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory*, 16(3). 425–458.
- Vassiliadou, Hélène, Francine Gerhard-Krait, Georgia Fotiadou & Lammert, Marie. 2023. *Pseudo(-)* in French and Greek: Categorization and approximation. *Zeitschrift für Wortbildung/Journal of Word Formation* 7(1). 234–262.
- Zeldes, Amir. 2012. *Productivity in argument selection: From morphology to syntax*. Berlin & New York: De Gruyter.

Ivan Lacić
Alma Mater Studiorum – University of Bologna
Department of Classical Philology and Italian Studies
Via Zamboni, 32
40126 Bologna, Italy
ivan.lacic2@unibo.it



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

Jan Seifert

Nutzende – Kritisierende – Forschende:

Zum Gebrauch substantivierter Partizipien als genderinklusive Referenzstrategie

Abstract: Pluralformen substantivierter Partizipien zur Personenbezeichnung wie *Teilnehmende* sind genusneutral und werden deshalb zum genderinklusive Sprachgebrauch empfohlen und genutzt. Es stellt sich jedoch die Frage, ob diese Formen semantisch äquivalent sind mit abgeleiteten Nomina agentis wie *Teilnehmer* und ob ihr Gebrauch mit den Regularitäten der deutschen Grammatik kompatibel ist. Umstritten ist insbesondere, ob ihnen eine spezifische aspektuelle Bedeutung zukommt. Zur Klärung dieser Fragen werden in der vorliegenden Studie synchrone und diachrone Daten aus zwei Zeitungskorpora (*die tageszeitung*, Jahrgänge 2000 und 2022, sowie Mannheimer Korpus Historischer Zeitungen bzw. Zeitschriften 1737–1905) erhoben und quantitativ wie qualitativ analysiert. Der Vergleich zeigt, dass die Partizipien semantisch weitgehend äquivalent mit deverbalen Nomina agentis sind und sie überwiegend zum Ausdruck von Eigenschaften, Gruppenzugehörigkeit und Gewohnheiten gebraucht werden. Festzustellen ist zudem, dass das Muster bereits im 19. Jahrhundert genutzt wurde und keineswegs eine neue Entwicklung ist.

Keywords: Personenbezeichnung, Partizip, Konversion, Nomen agentis, genderinklusive Sprache, Genderlinguistik

1. Einleitung: Partizipkonvertate als Referenzausdrücke

Im Zuge der Bemühungen um einen gendersensiblen Sprachgebrauch werden zur Personenreferenz verschiedene Arten von Ausdrücken empfohlen und genutzt, die sich hinsichtlich der morphologischen Struktur, des zugrunde liegenden Gender-Konzepts und der damit verbundenen pragmatischen Effekte und „Registrierungen“ (Kotthoff 2020) unterscheiden. Zur Vermeidung des geschlechtsübergreifenden Maskulinums kann im Sinne einer ‚Sichtbarmachung‘ explizit auf männliche und weibliche Personen verwiesen werden (z. B. *Studentinnen und Studenten*, *LehrerInnen*, *Dozent/-innen*), andere Formen sollen auf weitere Geschlechtsidentitäten Bezug nehmen (z. B. *Student*innen*, *Lehrer:innen*). Eine weitere Möglichkeit besteht darin, vom Genus zu abstrahieren: Pluralformen der Konversionen von Adjektiven und Partizipien (z. B. *Studierende*, *Lehrende*; *Angestellte*) bieten gegenüber den anderen genannten Formen den Vorteil,

genusneutral zu sein, sodass kein direkter Zusammenhang mit einem Sexus hergestellt werden kann (vgl. Dieward & Steinhauer 2017: 54; Kürschner 2019; Bross 2023; Kotthoff & Nübling 2024: 175).

Eine Zufallsauswahl alltagssprachlicher Beispiele aus unterschiedlichen Domänen zeigt, wie Substantivierungen des Partizip I in diesem Sinne verwendet werden. Dabei kommen native oder entlehnte Basen in Betracht, es kann sich um Simplicia oder um morphologisch komplexe (hier: Partikel-) Verben handeln, und der Grad der Referenzialität (vgl. Kotthoff & Nübling 2024: 105) ist unterschiedlich ausgeprägt:

- (1) Seit einer Woche läuft das Digitalsemester an Berliner Hochschulen – die Lehre findet wegen der Coronakrise ausschließlich online statt. Wir haben **Studierende** und **Dozierende** gefragt, was dabei klappt – und was nicht. [...] Zwar kann es nun nicht mehr passieren, dass – wie noch im vergangenen Semester – **Teilnehmende** aufgrund überbelegter Kurse auf dem Boden sitzen müssen, weil weder genug Stühle noch ausreichend PCs zur Verfügung stehen. Allerdings besitzen auch nicht alle **Studierenden** einen leistungsfähigen Laptop, der die Teilnahme an Zoom-Konferenzen uneingeschränkt ermöglicht.
(www.tagesspiegel.de/wissen/studierende-lesen-weniger-aber-kommunizieren-mehr-das-klappt-im-digitalsemester-an-den-berliner-unis-und-das-nicht/25784174.html) (aufgerufen am 8. März 2025)
- (2) Die **Forschenden** betonen, dass eine Verbesserung des Zugangs zu Grünflächen eine wichtige Strategie für die öffentliche Gesundheit und insbesondere zur Verringerung der **Rauchendenquote** darstellen kann.
(www.heilpraxisnet.de/naturheilpraxis/rauchen-aufhoeren-gruener-lebensraum-macht-es-menschen-leichter-20201102527217/) (aufgerufen am 8. März 2025)

Harnisch (2016: 159) sieht in diesem Verfahren im Prinzip eine „probate Strategie zur Vermeidung sowohl von generischen Formen (*Studenten*) als auch von Splitting-Syntagmen (*Studentinnen und Studenten*)“. Ein Vorteil dürfte auch darin bestehen, dass solche Formen keine Festlegung auf ein bestimmtes Verständnis von ‚Geschlecht‘ erfordern: Ausdrücke wie *Studentinnen und Studenten* oder *StudentInnen* entsprechen der „feministisch inspirierte[n] personenreferentielle[n] Praxis“ (Kotthoff 2020: 106), der ein dichotomisches Geschlechterverständnis zugrunde liegt, während Ausdrücke wie *Student*innen* dem „queeren Register“ zuzuordnen sind (Kotthoff 2020: 114) und auf ein nichtbinäres Geschlechterkonzept verweisen (sollen) (vgl. Ferstl & Nübling 2024: 262). Im Gegensatz dazu wird mit konvertierten Partizipien von der Geschlechtskategorie abstrahiert. Die tatsächliche Wirkung

derartiger Formen mit Differenzialgenus ist bislang nicht vollständig geklärt (vgl. Kotthoff & Nübling 2024: 140–141).¹ Sie scheinen mittlerweile aber so frequent (vgl. auch Sökefeld 2021: 121), dass sie bereits einen Reanalyseprozess durchlaufen und eine pragmatische Funktion bekommen haben: Der morphologische Marker *-end* ist, da die Sexus-Neutralisierung intendiert ist, zu einem „für politisch korrekte Ausdrucksweise genutzten pragmatischen Signal“ geworden (Harnisch 2016: 169) – mit dem Effekt, dass auch generisch maskuline Partizipialformen im Singular zu beobachten sind:

- (3) a. Auf Anfrage erhält **der Studierende** Auskunft über den Stand **seiner** Leistungspunkte.
 b. die Dankesrede, die **der Preistragende** hält
 (diese und weitere Belege in Harnisch 2016: 160 u. 164)

Angenommen wird üblicherweise, dass die genannten Konversionsprodukte semantisch prinzipiell äquivalent mit Nomina agentis auf *-er*, *-ent*, *-ant* usw. seien.² Der folgende Textauszug illustriert den Gebrauch weiterer morphosyntaktischer Ausdrucksformen, die als Personenbezeichnung genutzt werden – partizipiales Adjektiv, Konversionsprodukt, Nomen agentis (als Erstglied eines Kompositums) sowie Relativsatz:

- (4) Bei den 18- bis 25-Jährigen ist die **Raucherquote** deutlich höher als bei den jüngeren Jugendlichen. [...] Der Einstieg in das Rauchen wird gefördert durch: / **rauchende** Eltern / den Einfluss gleichaltriger Freunde, **die rauchen**. [...] Rund zwei Drittel **der Rauchenden** zwischen 12 und 25 Jahren möchten aufhören zu rauchen. (Hervorhebungen v. J.S.)
 (<https://rauchfrei-info.de/informieren/verbreitung-des-rauchens/raucherquote-bei-kindern-jugendlichen-1/>) (aufgerufen am 8. September 2021)

Es scheint so, als seien die in (4) gebrauchten Referenzausdrücke für ‚Personen, die rauchen‘ prinzipiell ohne Änderung der Semantik austauschbar gegen andere Formen, wie sie exemplarisch in (5) aufgelistet sind:

- (5) a. Raucherquote / Rauchendenquote
 b. rauchende Eltern, Freunde / Eltern, Freunde, die rauchen / Eltern, Freunde, die Raucher sind / Rauchereltern, Raucherfreunde
 c. Rauchende / Raucher / Raucherinnen und Raucher / Menschen, die rauchen

¹ Vgl. aber Kotthoff & Nübling (2024: 122–130) zu einzelnen experimentellen Studien, in denen die Verarbeitung von Neutralformen zumindest mitberücksichtigt wird, sowie Sato, Gabriel & Gyga (2016).

² So legen es die einschlägigen Empfehlungen nahe, vgl. beispielsweise Diewald & Steinhauer (2017: 54) oder schon früher das *Handbuch der Rechtsförmlichkeit* (HdR 2008: 52).

Der Unterschied bestünde dann nur in der formalen Realisierung bzw. der syntaktischen Einbindung.

Diese Auffassung ist nicht unumstritten; so mahnt Eisenberg (2020b: 17) an, „in jedem Einzelfall die Frage“ zu stellen, „ob die Partizipien dieselbe Bedeutung haben wie das Substantiv, das sie ersetzen sollen“; ein pauschaler Ersatz der Nomina agentis durch Partizipialformen sei ein „Eingriff in die Grammatik“. Das substantivierte Partizip, so führt er an anderer Stelle (Eisenberg 2021) am Beispiel von *Lesender* und *Sprechender* aus, habe eine engere Bedeutung: „[E]s geht nicht um die Tätigkeit als solche, sondern um eine aspektuelle Überformung derselben. Die Tätigkeit befindet sich im Verlauf, sie ist unabgeschlossen und in aller Regel an kontextuell gegebene Gleichzeitigkeit gebunden.“ Dass die Partizipien eine Gleichzeitigkeitslesart evozieren, wird in der populären Sprachkritik als Argument gegen den Gebrauch dieser Formen angeführt. Dabei wird häufig mit konstruierten Beispielen operiert, die die vermeintliche Widersinnigkeit konvertierter Partizipien illustrieren sollen:

Studierende sind Studenten, die gerade am Studieren sind. Wenn sie auf einer Demo protestieren, werden sie zu Protestierenden. Die Formulierung *protestierende Studierende* enthält daher einen Widerspruch, denn wenn Studenten protestieren, können sie nicht zur gleichen Zeit studieren. Ebenso widersinnig sind *schlafende Lehrende*, *Fahrrad fahrende Dozierende* oder *ermordete Teilnehmende*. (Payr 2022: 97)

Demgegenüber machen Diwald & Steinhauer (2017: 55) geltend, dass sich Partizipien wie Adjektive zur Bezeichnung von Zuständen oder inhärenten Eigenschaften eignen: „*Hungernde* können zwischendurch auch einmal gesättigt sein; *Reisende* können sich zeitweise an einer Stelle aufhalten – und *Studierende* sind eben auch Studierende, wenn sie gerade im Kino sind oder schlafen, weil sie grundsätzlich studieren.“

Letzteres wird durch den Befund bestätigt, dass die Bezeichnung *Studierende* seit dem 16. Jahrhundert etabliert ist, seit der ersten Hälfte des 20. Jahrhunderts auch als „bequemes commune in der amtssprache zur gattungsbezeichnung“ anstelle von *Studenten und Studentinnen* (DBW 20 [1942], 283; vgl. Glück 2020: 397–399). Das korrespondiert mit dem Ergebnis einer repräsentativen Umfrage zu Spracheinstellungen, in der auch nach der Bewertung von Personenbezeichnungen gefragt wurde. Demnach stößt die Variante *Studierende* mit 46 % Zustimmung auf die größte Zustimmung vor Splittingformen (*Studentinnen und Studenten*) und dem generischen Maskulinum (*Studenten*), die jeweils von 17 % der Befragten favorisiert werden (vgl. Adler & Plewnia 2019: 149–150). Ob ein

weniger lexikalisierten Referenzausdruck (z. B. *Forschende*) ähnlich gut beurteilt würde, steht dahin; es bleibt also offen, ob hier der Formtyp oder ein bestimmtes Lexem bewertet wurde.

Auch sonst stößt man bei der Lektüre älterer Texte durchaus auf Fälle, in denen substantivierte Partizipien anstelle von Nomina agentis wie *Hörer* oder *Photograph* gebraucht werden; die Belege (6) und (7) enthalten solche Zufallsfunde:

- (6) Zû zyten ist ouch hüpsch zereden vnd **den hörenden** angenäm wenn einr nach eim volkomen sinn der red: ettwas hinnach redt das denselben sinn sterckt [...]
(Knappe, Joachim & Stefanie Luppold (Hrsg.). 2009. *Friedrich Riederer. Spiegel der wahren Rhetorik (1493)*. Wiesbaden: Harrassowitz (Gratia 45), 126)
- (7) Wie für jeden **Photographierenden**, so hat auch für den Stereoskopiker die Frage [...] Bedeutung [...]. (Wolf-Czapek, K. W. 1914: *Stereoskopie*. Dresden: Ica, 18)

Es stellt sich nun die Frage, wie jenseits mehr oder weniger lexikalisierten (Einzel-?)Fälle das Verhältnis der Partizip-Konversionen zu deverbalen Nomina agentis zu beschreiben ist. Haben die Ausdrücke *Hörende*, *Lesende*, *Forschende*, *Teilnehmende* und *Rauchende* tatsächlich eine andere Semantik als *Hörer*, *Leser*, *Forscher*, *Teilnehmer*, *Raucher* bzw. die zugehörigen movierten Formen? Falls ja: Wie lässt sich die Wortbildungsbedeutung der beiden Typen systematisch beschreiben? Welche Bedingungen gelten jeweils, und welchen Restriktionen unterliegt die Bildung? In diachroner Hinsicht ist überdies zu klären, ob und inwiefern sich Regularitäten und Gebrauch innerhalb des Neuhochdeutschen geändert haben.

Diesen Fragen wird im vorliegenden Beitrag im Rahmen einer Korpusstudie nachgegangen. In Abschnitt 1 wird festgestellt, wie Bildung, Semantik und Gebrauch des Partizip I bzw. der entsprechenden Konversionsprodukte in Grammatiken bzw. der einschlägigen Forschungsliteratur beschrieben werden. Auf Basis dieser Auskünfte erfolgt ein systematischer Vergleich der beiden Bildungsmuster (Abschnitt 2). In Abschnitt 3 wird zunächst die Anlage der Korpusstudie erläutert (Abschnitt 3.1), ehe die Ergebnisse der synchron orientierten (Abschnitt 3.2) und der diachron orientierten Untersuchung (Abschnitt 3.4) präsentiert und diskutiert werden. Ziel ist es, jenseits konstruierter Einzelbeispiele ein differenziertes Bild der grammatischen Verhältnisse und vom tatsächlichen Gebrauch von Partizip-I-Konversionen zu gewinnen.

2. Das Partizip I – Status und Funktionen

2.1 Grammatikographischer Befund

Das Partizip I wird in Grammatiken unter drei Aspekten beschrieben: Erörtert werden einerseits morphologischer Status und syntaktisches Verhalten (und damit die Wortartenzugehörigkeit), andererseits die damit verbundene Semantik. Die Beschreibungen fallen keineswegs einhellig aus.

2.1.1 Morphologischer Status und syntaktisches Verhalten

Die IDS-Grammatik (Zifonun, Hoffmann & Strecker 1997: 2205–2207) plädiert dafür, das Partizip I der Klasse Adjektiv zuzuordnen; es sind „durch Wortbildung aus Verben entstandene Adjektive“ (2205), wobei „der Übergang in den Adjektivstatus nicht auf der Wort-, sondern auf der Phrasenebene stattfindet“, nämlich mittels des phrasalen Affixes {d}, was das Rektionsverhalten dieser Adjektive erklärt (vgl. 2207). Eine vergleichbare Auffassung scheint zunächst auch Eisenberg (2020a: 211) zu vertreten, wenn er vorschlägt, „das Part1 aus dem verbalen Paradigma herauszunehmen und es als deverbales Adjektiv anzusehen.“ An späterer Stelle (2020a: 269) verweist er jedoch auf die Nähe zum Verb, die sich „in der strikten Vererbung verbaler Argumente auf das Part1“ zeige (vgl. auch Zifonun, Hoffmann & Strecker 1997: 2206). Für die Duden-Grammatik (2022: 671) gibt die Kombination „sowohl verbale[r] als auch adjektivische[r] Eigenschaften“ dazu Anlass, von „Mischformen zwischen Adjektiv und Verb“ zu sprechen“. Sie verweist darauf, dass „auch bei attributiver Verwendung [...] die Partizipien viele ihrer verbalen Eigenschaften“ beibehielten (2022: 663); genannt werden jedoch nur syntaktische Merkmale, nicht aber semantische, was für die adäquate Beschreibung der Partizipien durchaus relevant wäre (s. unten). Genauer fassen es Fuhrhop & Teuber (2000: 112–113):

Das Partizip 1 übernimmt die Funktion, infinite Verben in attributiver und adverbialer Satzgliedrelation zu ermöglichen. Es tritt nie in regierter oder prädikativer Position auf und auch nicht als Teil einer Periphrase. Dort, wo das Partizip 1 steht, stehen typischerweise Adjektive; attributiv muß dann adjektivisch flektiert werden. Deshalb ist das Partizip 1 ‚adjektivisch‘.

In der IDS-Grammatik werden trotz der oben referierten Auffassung vom adjektivischen Status der Partizipien verschiedene Ausprägungsgrade des Adjektivstatus angesetzt. Kriterium dafür ist die Stellenzahl (*die mich faszinierende Angelegenheit* vs. *die*

faszinierende Angelegenheit; Zifonun, Hoffmann & Strecker 1997: 2207); unterschieden wird also zwischen einer „mehrstellig-phrasalen“ und einer „nullstelligen“ Variante, wobei letztere als Adjektiv klassifiziert werden soll. Der Status kann durch morphologische und syntaktische Testverfahren ermittelt werden, wobei zu prüfen ist, ob

- das Partizip prädikativ gebraucht werden kann (*Peter ist wütend* / **schlafend*)
- die Präfigierung mit *un-* möglich ist (*unpassend* / **unschlafend*)
- die Graduierung möglich ist (*wütender* / **schlafender*)

(Zifonun, Hoffmann & Strecker 1997: 2208–2209)

2.1.2 Semantik

Von besonderem Interesse ist zweifellos die Frage, welche Semantik dem Partizip zukommt und welche Konsequenzen sich daraus für das Konversionsprodukt ergeben. Da es sich um ein Wortbildungsprodukt handelt, müsste auch eine spezifische Wortbildungsbedeutung angegeben werden können, die dann auch bei der Konversion des Adjektivs zur substantivischen Personenbezeichnung beibehalten wird.

Glück (2020: 408) kommt nach einer Durchsicht verschiedener Grammatiken zu dem Ergebnis, dass die inhärente Bedeutung des Partizips I darin bestehe, „‘gleichzeitige Handlung im Vollzug, gleichzeitiges Geschehen im Ablauf’ in Bezug auf das finite Verb“ auszudrücken. Im Detail zeigen sich allerdings Unterschiede in der Bewertung und Gewichtung, wie nachfolgend skizziert wird.

In der Akademie-Grammatik wird zur Beschreibung der Semantik auf die Aspektkategorie Bezug genommen; das Partizip „kennzeichnet ein Geschehen als in seinem Verlauf unbegrenzt (durativ, imperfektiv)“ (Heidolph et al. 1981: 567). Dieser Auffassung folgt auch Eisenberg (2020a: 269; Hervorh. i. Orig.):

Die Bedeutung des Part1 kann aspektuell als imperfektiv und auf das Subjekt des vererbenden Verbs bezogen gekennzeichnet werden (**eine lesende Dichterin zu Eine Dichterin liest**): Der vom Verb bezeichnete Vorgang ist unabgeschlossen, er befindet sich im Verlauf.

Damit kann perfektiven bzw. telischen Verben eine imperfektive Form zur Seite gestellt werden; Partizipien wie *aufwachend*, *einschlafend*, *sterbend*, *aufweckend*, *tötend*; *entstehend*, *erblühend* usw. (vgl. die Beispiele für telische Verben in Duden 2022: 658) lassen eine Deutung im Sinne Eisenbergs zu.

In anderen Darstellungen wird das Partizip I hinsichtlich der zeitlichen Relation zu einem Referenzpunkt beschrieben. Nach Paul (1920: 68, § 318) markiert das Partizip „Gleichzeitigkeit“ mit dem „Zeitpunkt“ – nicht einer Zeitspanne –, der durch das finite Verb festgelegt wird. Ähnlich verfährt die Duden-Grammatik; sie scheint sich von dieser Einschätzung allerdings sukzessive zu distanzieren. Hieß es zunächst: „Wenn nichts dagegen spricht, drückt es Gleichzeitigkeit mit dem übergeordneten Geschehen aus. Es kann aber auch Gegenwartsbezug ausdrücken oder separat durch ein Temporaladverbiale bestimmt werden [...]“ (Duden 2009: 563), so wurde das in der Folgeauflage (2016: 577) etwas eingeschränkt (Hervorh. v. J.S.):

Meistens wird das vom attributiven Partizip I [...] beschriebene Geschehen als gleichzeitig mit dem übergeordneten Geschehen **verstanden** [...]. Es kann jedoch auch Gegenwartsbezug ausdrücken [...] oder separat durch ein Temporaladverbiale bestimmt werden [...].

Bemerkenswerterweise finden sich in der jüngsten Auflage (Duden 2022) in den einschlägigen Abschnitten keinerlei Angaben zur spezifischen Bedeutung des Partizips, lediglich der Hinweis, dass die Konversionsprodukte sehr häufig als Personenbezeichnungen gebraucht werden (Duden 2022: 726, vgl. ebd. 702).

Für die IDS-Grammatik hat die Ausprägung des Adjektivstatus bzw. der zunehmende Lexikalisierungsgrad Konsequenzen für die semantische Beschreibung: „Volladjektivische Partizipien I abstrahieren von dem von der entsprechenden Verbgruppe bezeichneten Ereignis; das heißt, Gegenstände werden nicht im Hinblick auf Situationen charakterisiert, sondern im Hinblick auf situationsabgelöste, verallgemeinerte Eigenschaften oder Dispositionen“ (Zifonun, Hoffmann & Strecker 1997: 2208–2209); als Beispiele werden *reizend* und *bedeutend* genannt.

2.1.3 Konversionsprodukte

Die referierten Darstellungen beziehen sich zunächst nur das Partizip I in seiner Eigenschaft als Adjektiv. Bei den Personenbezeichnungen, um die es hier geht (vgl. Duden 2022: 702), handelt es sich aber um Konversionsprodukte, und es stellt sich die Frage, wie das Verhältnis dieser Substantive zum unmittelbar zugrundeliegenden Partizip und zur verbalen Basis zu beschreiben ist. Bei der syntaktischen Konversion handelt es sich per definitionem um eine reine Umkategorisierung der Wortart gemäß (8):

- (8) a. rauchende Personen
- b. Rauchende

Dennoch muss die Gemeinsamkeit syntaktischer Kontexte in Rechnung gestellt werden.

Eisenberg (2020a: 306) weist darauf hin, dass

[e]ine Folge dieser ›syntaktischen Überlappung‹ von Adjektiv und substantiviertem Adjektiv [darin besteht], dass letzteres auf ersteres bezogen bleibt. Jedes Adjektiv ist [...] substantivierbar und es kann als Substantivierung lexikalisiert und sogar idiomatisiert werden, aber es behält immer auch seine Bedeutung als transparentes Konversionsprodukt.

Er nennt in diesem Zusammenhang auch Partizip-Konversionen wie *entscheidend* – *der/die/das Entscheidende* (2020a: 305) und *die abgeordnete/Abgeordnete* (2020a: 306). Es liegt nahe, dass die oben beschriebenen grammatischen und semantischen Eigenschaften der Partizipien erhalten bleiben, und das hat Konsequenzen für die syntaktische Einbindung, wie noch darzulegen ist (s. Abschnitt 2.3.2).

Die einfachste Paraphrase departizipialer Personenbezeichnungen des Typs (8b.) (*Rauchende, Teilnehmende, Studierende*) lautet ‚Personen, die rauchen/teilnehmen/studieren‘. Im Falle volladjektivischer Partizipien wäre gem. Zifonun, Hoffmann & Strecker (1997: 2008–2009) eine abstrakte Klassifikation anzusetzen wie ‚durch die Eigenschaft X gekennzeichnet‘.

Ob es sich um ‚Eigenschaften‘ oder ‚Dispositionen‘ handelt, dürfte von der Semantik des Basisverbs abhängen. Es greift auch grundsätzlich zu kurz, Verben pauschal als Ausdrucksmittel für ‚Handlungen‘ oder ein ‚Geschehen‘ aufzufassen. Es bietet sich vielmehr an, mit v. Polenz (2008: 159–167) verschiedene **Prädikatsklassen** anzusetzen, mit denen die Bedeutung systematisch erfasst werden kann; zu unterscheiden sind u.a. Handlungs-, Vorgangs-, Eigenschafts- und Zustandsverben. Handlungen sind als intentionale Tätigkeiten zu beschreiben, zu denen aufgefördert werden kann, die auch unterlassen werden können und mit denen eine Zuschreibung von Verantwortlichkeit verbunden ist. Demgegenüber sind Vorgänge dadurch bestimmt, dass ein Geschehen sich unabsichtlich ohne Einwirkung einer handelnden Instanz vollzieht. Die Abgrenzung ist, beispielsweise bei *verba sentiendi*, nur kontextbezogen möglich (v. Polenz 2008: 161). Mit Zustandsverben werden Prädikationen über „grundsätzlich veränderliche physische oder psychische Zustände“ vollzogen, zu denen auch Einstellungen gehören; Beispiele sind *leben, wissen, sich bekennen, leugnen*. Eigenschaftsprädikate schließlich betreffen

unveränderliche, dauerhafte Merkmale, wobei die Dauerhaftigkeit insbesondere „bei sozialen Eigenschaften als Status- oder Rollenbegriffe“ relativ ist:

Jemand, der an einem Abend oder auf einer Reise *nicht raucht*, ist noch kein *Nichtraucher*, wohl aber einer, der dieses Verhalten mit festem Vorsatz für [!] weitere Zukunft schon einige Wochen durchgehalten hat. *Beamter* ist man für Jahrzehnte, *Student* für Jahre, *Kandidat* für Monate oder Tage, *Diskussionsleiter* für Stunden oder Minuten.“ (v. Polenz 2008: 163).

Wenn Diwald & Steinhauer (2017: 55) feststellen, dass „*Studierende* [...] auch Studierende [sind], wenn sie gerade im Kino sind oder schlafen, weil sie grundsätzlich studieren“, zielen sie auf genau diese Unterscheidung von Prädikatsklassen ab: Dass diese Personen „grundsätzlich studieren“, bedeutet, dass sie einer bestimmten Personengruppe angehören, also einen bestimmten Status haben bzw. temporär durch eine bestimmte Eigenschaft gekennzeichnet sind. Das Verb *studieren* ist hier als Eigenschaftsprädikat zu klassifizieren. Sein Gebrauch als Handlungsverb dürfte hingegen unrealistisch bzw. auf bestimmte Kontexte beschränkt sein, wie (9) illustriert:

- (9) a. ?Ich bin in der Bibliothek und studiere gerade³
 b. Dann studier endlich mal und surf nicht so viel 😊
 (<https://forum.buschtaxi.org/unser-magnus-und-seine-hormone-t10784-255.html>)
 (aufgerufen am 8. März 2025)

Im beschriebenen Sinne können Verben je nach Verwendungsweise durchaus verschiedenen Prädikatsklassen zugeordnet werden, beispielsweise *sehen* als HANDLUNG (*auf die Uhr sehen*), TÄTIGKEIT (*sich müde sehen*), EINSTELLUNG (*hoffnungsvoll in die Zukunft sehen*), VORGANG (*sie hat ihn nicht gesehen*), ZUSTAND (*das Boot sah nur ein Stück aus dem Wasser*), EIGENSCHAFT (*er sieht schlecht*) (v. Polenz 2008: 166).

Für das in Rede stehende Verb *rauchen* ergeben sich zumindest die Prädikatsklassen HANDLUNG (10a) und EIGENSCHAFT („Gewohnheit“, „Laster“) (10b):

- (10) a. Auf einem Flug nach Australien hat ein Deutscher immer wieder versucht zu rauchen.
 (welt.de, 03.01.2014)
 b. Anna ist Architektin und raucht seit ihrer Jugend. (<https://einfachnichtraucherin.ac-page.com/einfach-nichtraucherin-gold-lp/>) (aufgerufen am 8. März 2025)

³ Anders verhält es sich, wenn *studieren* transitiv in der Bedeutung ‚etwas gründlich lesen/betrachten‘ gebraucht wird: „Man kann aber auch eine Speisekarte, Prozessakten oder das Mienenspiel seines Gegenübers studieren“ (Glück 2020: 399) – aber diese Verwendungsweise ist hier ja gerade nicht relevant.

Phrasen wie *eine rauchende Architektin* ist demnach ambig:

- ‚eine Architektin, die zu einem Referenzzeitpunkt raucht (geraucht hat)‘
- ‚eine Architektin, die gewohnheitsmäßig raucht‘

Gleiches gilt für Eisenbergs (2020a: 269) *lesende Dichterin* (‚eine Dichterin, die zu einem Referenzzeitpunkt liest/gelesen hat‘ bzw. ‚eine Dichterin, die regelmäßig/gern liest‘). Als Testverfahren bietet sich die Paraphrase mittels *am*-Progressiv an; diese Form dürfte im Fall des EIGENSCHAFTS-Prädikats ausgeschlossen sein, da die zeitliche bzw. aspektuelle Dimension nicht Teil der Wortbildungsbedeutung ist:

- (10) c. ??Anna ist Architektin und seit ihrer Jugend am Rauchen.

Inwieweit derartige Formen tatsächlich ausgeschlossen sind, müsste allerdings im Rahmen einer Akzeptabilitätsstudie überprüft werden.

Welche Prädikatsklassen anzusetzen sind, ist von Verb zu Verb unterschiedlich; so ist eine *streikende Lokführerin* eine, ‚die sich am Streik beteiligt‘ (ZUSTAND), aber kaum eine, ‚die gewohnheitsmäßig streikt‘. Gleiches gilt für die *essende/schlafende Lokführerin*: Dass die habituelle Lesart (EIGENSCHAFT) hier weniger in Betracht kommt, ist eine Frage der Relevanz (es sei denn, es soll ein Übermaß ausgedrückt werden: ‚Lokführerin, die ständig isst‘ usw.).

2.2 Konvertierte Partizipien vs. Nomina agentis

Wie oben festgestellt wurde, ist die Bedeutung departizipialer Konversionsprodukte nicht ohne Weiteres aufgrund ihrer Form vorhersagbar, vielmehr sind mehrere Lesarten möglich, die im Zusammenhang mit der Verbsemantik stehen. Diese Ambiguität ist allerdings kein exklusives Merkmal substantivierter Partizipien: Auch andere Wortbildungsmuster, die deverbale Nomina agentis hervorbringen, lassen mehr als eine Lesart zu. Das gilt auch für die produktiven Bildungen auf *-er* und andere Suffixe, mit denen die konvertierten Partizipien konkurrieren (vgl. Fleischer & Barz 2012: 201–202). Es scheint zunächst, als seien diese Ausdrücke weitgehend synonym zu gebrauchen. Fleischer & Barz (2012: 202) erkennen aber einen „semantische[n] Unterschied zwischen beiden Bildungsweisen“, der vor allem darin bestehe,

dass das *-er*-Derivat in stärkerem Maße zur Bezeichnung eines Personenbegriffs mit festen Merkmalen tendiert (vgl. die Untergruppen ‚professionell‘, ‚habituell‘), womit auch eine Demotivation verbunden sein kann [...]. Nur in den seltensten Fällen hat das konvertierte Partizip I einen ähnlichen semantischen Charakter: *der Vorsitzende* (woneben auch *Vorsitzer*), *Reisende*, *Streikende* (woneben keine *-er*-Derivate).

Diese Einschätzung ist insofern bemerkenswert, als hier zwei Problembereiche angesprochen sind, die für einen angemessenen Vergleich der Bildungsmuster relevant sind:

1. Die abweichende Semantik, 2. die Beschränkungen bei der Anwendung des Bildungsmusters.

Zu 1.: Die Bedeutung der deverbalen Nomina agentis auf *-er* mag von der Bedeutung der Verben stärker abstrahieren; vorhersagbar ist sie damit aber nicht unbedingt, wie der Hinweis auf die semantischen Untergruppen deutlich macht. Die von Fleischer & Barz vorgeschlagene Differenzierung deverbalen Personenbezeichnungen in [PROFESSIONELL], [HABITUELL] und [OKKASIONELL] kann mit Scherer (2005: 101) auf alle Personenbezeichnungen ausgedehnt werden, selbstverständlich auch auf die konvertierten Partizipien. Die professionelle Lesart ist dabei allerdings nur ein „Spezialfall der habituellen Lesart“ (Scherer 2005: 102). Ergänzend kommt die Lesart [AKTUELL] hinzu, bei der der Aspekt der zeitlichen Relation dominiert. Zwar haben Experimente mit der Interpretation von Gelegenheitsbildungen gezeigt, dass die habituelle Lesart den Default-Fall auch bei nichtverbalen *-er*-Derivaten markiert (vgl. Scherer 2005: 101–102); dennoch kann im Zusammenhang mit Nomina agentis nicht unbedingt von „festen Merkmalen“ (s. o.) gesprochen werden. Welche Bedeutung ein Nomen agentis hat, hängt vom Gebrauch ab. Der Bildung *Tänzer* beispielsweise kommen – je nach Situation und Kontext – verschiedene Lesarten zu:

1. ‚jemand, der tanzen kann und regelmäßig oder gelegentlich tanzt‘ (habituelle / okkasionelle Bedeutung)
2. ‚jemand, der den künstlerischen Tanz regelmäßig auf bestimmtem Niveau gegen Honorar ausübt‘ (professionelle Bedeutung)
3. ‚jemand, der sich zu einem bestimmten Zeitpunkt auf einer Tanzfläche aufhält‘ (aktuelle Bedeutung)

Wer ein passionierter *Tänzer* ist, übt diese Tätigkeit nicht zwingend professionell aus, und er muss sich zum Referenzzeitpunkt auch nicht auf einer Tanzfläche aufhalten; wer sich zu einem bestimmten Anlass vielleicht einmal – vielleicht mehr schlecht als recht – zur

Musik bewegt, ist dann *Tänzer*, verdient damit aber weder Geld, noch muss er dieser Tätigkeit regelmäßig nachgehen.

Im Fall von *Leser* liegt eine habituelle Lesart nahe („wer gern/regelmäßig liest“; vgl. *Nichtleser*); je nach Situation ist aber die aktuelle Lesart (hier Rolle: „Rezipient“) relevant.

In (11) wiederum ist für *Esser* die aktuelle Lesart relevant („wer zum Zeitpunkt x isst“):

- (11) Aber – keine Sorge: moderne Sensorik bedeutet vor allem, die Praxis des Essers und ein gutes Geschmackserlebnis zu antizipieren und keineswegs eine Komplizierung des Essens (wie es manchmal vermutet wird). (F.A.Z., 25.07.16)

Die aktuelle Lesart kann aber durch den Kontext auch ausgeschlossen sein, wie (12a) illustriert:

- (12) a. Die meisten Autofahrer sind nämlich auch Fahrradfahrer und umgekehrt. (Leserbrief, GA Bonn, 10.03.22)
b. ??Die meisten Autofahrenden sind nämlich auch Fahrradfahrende und umgekehrt.

In (12a) ist von einer habituellen oder okkasionellen Lesart auszugehen, weil sich sonst ein logischer Widerspruch ergäbe. Ob (12b) akzeptabel ist oder nicht, hängt von der Lesart ab; wenn mit dem konvertierten Partizip eine aspektuelle und damit aktuelle Lesart verbunden ist, müsste der Satz zurückgewiesen werden (vgl. Abschnitt 2.3.1).

Zu 2.: Auch wenn das *-er*-Bildungsmuster als hochgradig produktiv gelten kann, unterliegt es Beschränkungen unterschiedlicher Art (vgl. ausführlich Scherer 2005: 91–94). Hier zeigt sich ein grundsätzlicher Unterschied zu den Partizipkonvertaten, die im Prinzip „universal“ bildbar sind (Fleischer & Barz 2012: 202). Dass bei Konversionsprodukten, anders als bei Derivaten, keine Blockierungen auftreten, ist nach Wurzel (1988: 197–198) darauf zurückzuführen, dass sie „keine spezifische Kategoriensemantik besitzen“, sondern sich „wie semantikarme Flexionsformen verhalten.“

Das erklärt aber nicht alle Fälle der Blockierung von Derivationsmustern: Wenn nach Wurzel (1988: 183) ein Derivationsmuster „aufgrund der Existenz von synonymen Bildungen anderer Form blockiert“ ist – er führt hier das Beispiel **Reiser* („wegen *Reisender*“) an –, stellt sich die Frage, warum sich das substantivierte Partizip überhaupt hat etablieren können und andererseits die im 18. Jh. belegte Form *Reiser* (DWB VIII: 739) untergegangen ist.

Die Fälle (13) und (14) illustrieren Beschränkungen der *er*-Derivation im Zusammenhang mit Fremdwörtern:

- (13) Streikende / *Streiker (Fleischer & Barz 2012: 202)
- (14) a. Demonstrierende / Demonstranten / ?Demonstrierer
b. Protestierende / #Protestanten / ?Protestierer

Man könnte vermuten, das konvertierte Partizip trete an die Stelle von Nomina agentis wie ?*Streiker*, ?*Demonstrierer* oder ?*Protestierer*, weil deren Bildung ausgeschlossen sei. Abgesehen davon, dass der Hinweis auf die Existenz synonyme Bildungen eine zirkuläre Erklärung ist, lassen sich aber durchaus zahlreiche Belege für entsprechende *-er*-Derivate finden⁴; ausgeschlossen sind die Formen also keineswegs. Darüber hinaus existieren neben den Partizipien weitere Bildungsmuster (vgl. Fleischer & Barz 2012: 245); in (14) wäre *Demonstranten* eine mögliche Variante, während in (14b) *Protestanten* wegen abweichender Semantik nicht in Betracht kommt. Bei der Bildung eines Nomen agentis zum Verb *photographieren* (s. oben Beispiel 7) ist *Photographierende* keineswegs zwingend; die Reduktion auf das Konfix *-graph* (*Photograph*) ist seit 1859 im Deutschen belegt und etabliert (vgl. DFWB 5 [2004]: 1061). Anzunehmen ist allerdings, dass dieses Substantiv zu Beginn des 20. Jh.s die professionelle Lesart evoziert hat (vgl. die ebd. genannten Belege) und mit dem Partizip eine ‚allgemeinere‘ Bedeutung verbunden ist. Das wiederum erklärt freilich nicht, warum in (6) *Hörende* gebraucht wird, da *Hörer* seit dem Fnhd. etabliert ist. Hier ist der Entstehungskontext zu berücksichtigen, denn der fragliche Text, Riederers Rhetorik, basiert auf verschiedenen lateinischen Quellen – es ist also damit zu rechnen, dass hier ein etabliertes lateinisches Strukturmuster (*audientes* ‚Hörende‘ > ‚Hörer‘) ins Deutsche übertragen wurde.

Schon diese beiden Belege zeigen eindrücklich, dass für eine angemessene Beschreibung der beiden Bildungsmuster auch die historische Perspektive zu berücksichtigen ist (s. Abschnitt 3.4). Insgesamt ergibt sich aus den Überlegungen, dass die beiden Wortbildungsverfahren als eigenständige und konkurrierende Möglichkeiten der Bildung von Personenbezeichnungen zu betrachten sind, sie stehen also nicht in einem komplementären Verhältnis.

⁴ Eine DeReKo-Abfrage (Archiv der geschriebenen Sprache W-öffentlich) erbringt für *Demonstrierer* 106, für *Streiker* 108 und für *Protestierer* sogar 4.054 Treffer.

2.3 Beschränkungen beim Gebrauch der Partizipien

2.3.1 Zeitliche Relation

Klärungsbedürftig ist die bereits angesprochene zeitliche Relation zwischen dem vom Partizip Bezeichneten und dem „übergeordneten Geschehen“ (so z. B. Duden 2016: 577).

Das Bemühen, ein generisches Maskulinum wie *Absolventen* – etymologisch ebenfalls ein Part. I – zu vermeiden, führt in (15) zum Gebrauch von *Absolvierende*.

- (15) Das primäre Arbeitsfeld für **Absolvierende** des Masterstudiengangs Medical and Health Education ist die Tätigkeit als wissenschaftliche:r Mitarbeiter:in mit Lehrverpflichtung bzw. die Amtsausübung als Professor:in an Universitäten und Hochschulen im Bereich der Medizin und Gesundheit.
(<https://www.cene-nachwuchsfoerderung.de/qualifikationsphase-postdoc-und-junior-professur/medical-and-health-education>) (aufgerufen am 8. März 2025)

Intendiert ist offenbar eine Vorzeitigkeitslesart im Sinne von ‚die den Masterstudiengang absolviert [und den Abschluss erworben] haben‘. Wird Gleichzeitigkeit mit dem übergeordneten Geschehen – der Beschäftigung an Hochschulen – angesetzt, käme eine Lesart ‚die dabei sind, den Abschluss zu erwerben‘ in Betracht, die vom Kontext allerdings nicht gedeckt ist.

Auch für die Fälle in (16) ist eine Gleichzeitigkeitsrelation wenig wahrscheinlich, wenn nicht sachlich ausgeschlossen.

- (16) Gewinnende, Verlierende / Gewinner, Verlierer

Gewinner bzw. Verlierer ist, wer gewinnt/verliert oder gewonnen/verloren hat; das Nomen agentis ist unmarkiert hinsichtlich der zeitlichen Relation. Beim Partizip 1 kommt, wie oben dargelegt wurde, allenfalls eine Gleichzeitigkeitsrelation in Betracht, nicht jedoch Nachzeitigkeit, weshalb (17b) zweifelhaft erscheint.

- (17) a. Als Weihnachtsgeschenk wurden unter allen Teilnehmern tolle Geschenke verlost. Die glücklichen Gewinner stehen nun fest.
(<https://www.daserste.de/unterhaltung/soaps-telenovelas/sturm-der-liebe/specials/adventskalender-2013-gewinner-100.html>) (aufgerufen am 8. März 2025)
b. ??Die glücklichen Gewinnenden stehen nun fest.

Dass sich Partizipien (zu überwiegend telischen Verben) und *-er*-Derivate hinsichtlich der ausgedrückten Zeitrelation unterscheiden, illustrieren auch die Beispiele in (18):

- (18) a. Gutachtende / Gutachter
b. Wählende (die an der Wahl teilnehmen bzw. den Stimmzettel ausfüllen) / Wähler (wer das Wahlrecht hat) (vgl. Glück 2020: 389)
c. Teilnehmende ‚wer gerade teilnimmt‘ / Teilnehmer
d. ?Kaufende / Käufer ‚wer kauft / gekauft hat‘

Wer *Teilnehmer* ist, nimmt aktuell an etwas teil, kann aber auch bereits teilgenommen haben; *Käufer* können etwas gekauft haben; *Wähler* besitzen das Wahlrecht; in (19) sind es aber Personen, die gewählt haben:

- (19) Viele von Trumps Wählern wollen ihm einfach glauben. Hörten sie damit auf, so liefe das auf ein Eingeständnis heraus, dass all die Jahre doch die anderen recht hatten. (F.A.Z., 23.06.22)

Wie stark die Akzeptabilität im Einzelfall vom jeweiligen Kontext abhängt, illustriert der folgende Beleg.

- (20) Es kann vorkommen, dass ein Autor, der für ausreichend diskurs- und mitteilungsfreudig gehalten wird, nicht nur zu Lesungen, sondern auch zu Vorträgen und Seminaren eingeladen wird. Er darf dort im Prinzip über alles reden, lieber aber ist es den **Einladenden**, wenn er „Zur Lage der Gegenwartsliteratur“ spricht oder über „Neueste Tendenzen zeitgenössischer Literatur“. (Hans Ulrich Treichel; taz 13.05.2000)

Gastgeber bzw. Veranstalter sind, da die Einladung der Veranstaltung vorausgeht, in aller Regel nicht *Einladende*, sondern gewissermaßen *eingeladen Habende*, es ist im Normalfall also eine Vorzeitigkeitsrelation anzunehmen. Im vorliegenden Fall allerdings wird offenbar tatsächlich der Zeitpunkt der Einladung (die ausgesprochen, aber noch gar nicht angenommen sein muss) in den Blick genommen; insofern dürfte die Gleichzeitigkeitslesart hier unproblematisch sein.

Die Beispiele illustrieren einen grundsätzlichen Unterschied zwischen *-er*-Bildungen und konvertierten Partizipien: Bei den *-er*-Bildungen wird von einer zeitlichen Relation abstrahiert, sie sind diesbezüglich unmarkiert. Die Konversionsprodukte hingegen scheinen eine Gleichzeitigkeitsrelation nahezulegen.

Ob mit den konvertierten Partizipien mental eine Gleichzeitigkeitsrelation verbunden wird, ist Gegenstand einer Pilotstudie von Bross (2023). Die Gewährspersonen sollten Sätze des Typs *Die Arbeitenden lagen am See* daraufhin beurteilen, ob sie einen Widerspruch enthielten oder nicht. Die Ergebnisse deuten darauf hin, dass eine

„Gleichzeitigkeitsinterpretation zwischen dem im Matrixverb ausgedrückten Ereignis und dem in einem nominalisierten Partizip ausgedrückten Ereignis“ dann zu erwarten ist, wenn „die Partizipien nicht lexikalisiert sind“ (Bross 2023: 43); das gilt für Ad-hoc-Partizipien wie *Autofahrende*, *Kochende*, *Lesende* usw. (vgl. Bross 2023: 41). Als unproblematisch erwiesen sich hingegen frequente und damit weitgehend lexikalisierte Ausdrücke wie *Studierende*, *Mitarbeitende*, *Forschende*, *Teilnehmende* usw. Bross nimmt daher an, dass bei hoher

Gebrauchsfrequenz und damit steigendem Lexikalisierungsgrad der Partizipkonvertate die Gleichzeitigkeitsrelation an Relevanz verliert. Klärungsbedürftig bleibt allerdings, unter welchen Bedingungen die Ausweitung des Gebrauchs im Einzelfall überhaupt möglich ist und welche Faktoren die Lexikalisierung determinieren können.

An dieser Stelle sei darauf hingewiesen, dass auch die oben diskutierten Probleme der Vorzeitigkeitsrelation, die sich im Zusammenhang mit Partizipien wie *Gewinnende* oder *Absolvierende* stellen, der empirischen Überprüfung bedürfen.

2.3.2 Unterschiede in der syntaktischen Einbindung

Partizipien zeigen insofern eine gewisse Verbnähe, als das Prinzip einer „strikten Vererbung verbaler Argumente auf das Part1“ gilt, und zwar „für alle Objektkasus wie für das Präpositionalobjekt“ (Eisenberg 2020b: 269; vgl. Duden 2022: 663). Das dürfte aber Konsequenzen für die Struktur der Nominalphrase haben; so stellt sich die Frage, ob die betreffenden Ergänzungen als Genitivattribute realisiert werden können. Das sei am Beispiel *Teilnehmende* im Vergleich mit dem Nomen agentis *Teilnehmer* illustriert (21):

- (21) a. Teilnehmer des Seminars
 b. ?Teilnehmende des Seminars
 c. ?Teilnehmende am Seminar
 d. am Seminar Teilnehmende
 e. am Seminar teilnehmende Personen
 f. ?Seminarteilnehmende

Während beim Nomen agentis das Präpositionalobjekt als Genitivattribut realisiert wird (21a), ist das beim konvertierten Partizip nicht so eindeutig. Ob allerdings ein Präpositionalattribut in Betracht kommt (21c), erscheint zumindest fraglich; wenn das Prinzip der

Argumentvererbung gilt, müsste die verbale Struktur erhalten bleiben, was eine Abfolge wie in (21d) nahelegt. In diesem Fall entspräche die Struktur der einer Nominalphrase mit erweitertem Partizipialattribut (21e), bei der die Erweiterungsglieder zentripetal angeordnet sind. Als Variante erscheint grundsätzlich auch ein Determinativ-kompositum gem. (21f) möglich. Die Akzeptabilität von Formen gemäß (21b und 21c) kann letztlich nur empirisch festgestellt werden, beispielsweise durch eine Informantenbefragung; die Grammatiken bieten zu dieser Frage keinen Aufschluss.⁵ Andererseits scheinen bei etablierten Partizipkonvertaten wie *Studierende* Genitivattribute unauffällig: *Studierende der Universität Bonn*; *Studierende der Germanistik* vs. *?Germanistik Studierende / Germanistikstudierende*; auch dies ist ein Hinweis darauf, dass die Konversionsprodukte mit zunehmendem Lexikalisierungsgrad ihre partizipialen Eigenschaften verlieren.

2.4 Zwischenfazit

Aus den vorgetragenen Überlegungen ergibt sich als vorläufiger Befund:

- Das semantische Potenzial des Partizip 1 umfasst aspektuelle Funktionen.
- Die aspektuelle Lesart ist nicht zwingend; sie hängt von der Semantik/Prädikatsklasse des Basisverbs ab.
- Konvertierte Partizipien konkurrieren prinzipiell mit deverbale Nomina agentis; ihre Bildung unterliegt nur geringen Beschränkungen.
- Ob für konvertierte Partizipien eine Gleichzeitigkeitslesart in Betracht kommt, hängt auch vom Lexikalisierungsgrad ab. Partizipien mit Vorzeitigkeitslesart sind belegt, erscheinen aber zweifelhaft und müssten im Rahmen einer Akzeptabilitätsstudie überprüft werden.
- Das morphosyntaktische Verhalten der Konversionsprodukte unterscheidet sich theoretisch von dem der *-er*-Derivate; genauere Daten liegen aber bislang nicht vor.

⁵ Eisenberg (2020b: 269) merkt weiterhin – nicht frei von Ironie – an: „Neben dem ›Geerbten‹ können Modifikatoren jeder Art auftreten, so dass vor allem im Geschriebenen leicht einmal eine das Verständnis des jeweils Gemeinten im Allgemeinen unnötig erschwerende Konstruktion entsteht“; welche Modifikationen das sind und ob sie nur ‚schwierig‘ oder grundsätzlich systeminadäquat sind, führt er nicht aus.

3. Korpusstudien

Ziel der nachfolgenden Untersuchung ist es, zu ermitteln, wie substantivierte Partizipien in standardsprachlichen Texten zur Personenreferenz genutzt werden. Dabei sind die folgenden Leitfragen relevant:

1. Welche substantivierten Partizipien, die zur Personenreferenz genutzt werden, sind belegt?
2. Inwiefern handelt es sich um lexikalisierte Formen?
3. In welchem Verhältnis stehen sie zu deverbale Nomina agentis? Welche aspektuellen und temporalen Lesarten sind im Vergleich mit deverbale Nomina agentis jeweils anzusetzen?
4. Welcher Status kommt den substantivierten Partizipien im Vergleich mit anderen gegenderten Personenbezeichnungen zu?
5. Inwiefern zeigen sich hinsichtlich der Befunde diachrone Veränderungen?

3.1 Korpora und Methodisches

Untersuchungsgrundlage bildet das Deutsche Referenzkorpus (DeReKo) des Leibniz-Instituts für deutsche Sprache, das über die Anwendung Cosmas II elektronisch zugänglich ist. Angesichts der Korpusgröße war eine Beschränkung auf ein Teilkorpus unumgänglich. Geschriebene Standardsprache wird üblicherweise mit dem Sprachgebrauch überregionaler Zeitungen assoziiert (vgl. Eisenberg 2007: 217; Klein 2013: 25); die Beschränkung auf ein Zeitungskorpus bietet sich auch deshalb an, weil eine gewisse Systematik im Sprachgebrauch zu erwarten ist. Die Wahl fiel auf die Berliner *tageszeitung* (taz), weil hier einerseits eine Affinität zu gegenderten Personenbezeichnungen gegeben ist (vgl. Waldendorf 2024) und andererseits auch eine große Bandbreite genutzter Formen festgestellt wurde (vgl. Sökefeld 2021: 122; Waldendorf 2024). Berücksichtigt wurden die Jahrgänge 2000 und 2022, wie noch darzulegen ist.

Was die postulierte Standardsprachlichkeit der Korpusgrundlage betrifft, sei darauf hingewiesen, dass die taz in ihrer politischen Ausrichtung und mit ihrem Sprachgebrauch, der durch Expressivität und ungewöhnliche Wortbildungen gekennzeichnet ist, keinesfalls als repräsentativ für die überregionale Presse gelten kann. Das Korpus kann also nur einen Ausschnitt aus der – ohnehin durch Variantenreichtum gekennzeichneten (vgl. Klein 2013: 26) – Standardsprache widerspiegeln. Für die Fragestellung des

vorliegenden Beitrags ist das aber insofern unproblematisch, als es darauf ankommt, möglichst viele Belege zu erfassen und zu beschreiben.

Für einen diachronen Vergleich wurden zusätzlich zwei historische Zeitungskorpora durchsucht, das Mannheimer Korpus Historischer Zeitungen 1737–1877 (KHZ) sowie das Mannheimer Korpus Historischer Zeitschriften 1834–1905 (KHM). Diese beiden Teilkorpora umfassen 652 Texte mit 4.093.636 Wortformen.

Da eine gezielte Suche nach substantivierten Partizipien nicht möglich ist, worauf das IDS in seiner Hilfestellung selbst hinweist⁶, bestand nur die Möglichkeit, sämtliche Wortformen zu ermitteln, die auf die Zeichenfolge <ende> bzw. <enden> auslauten und Großschreibung am Wortanfang aufweisen. Die entsprechende Suchanfrage lautete: `#REG(^[A-Z][a-z]+ende[n]?$)`. Auf diese Weise werden alle schwach flektierten Formen im Singular und Plural sowie starke Formen des Nom./Akk. Pl. berücksichtigt, auf die es im vorliegenden Kontext ankommt. Einschränkend ist zu sagen, dass starke Formen im Nom. Sing. (*ein Studierender*) und Gen. Pl. (*die Bedürfnisse Studierender*) damit zwangsläufig unberücksichtigt bleiben.

Für den Zeitraum 2000–2022 wurden in der taz für die fragliche Zeichenfolge zunächst insgesamt 365.046 Treffer ermittelt. Pro Jahrgang sind es ca. 16.500 Fälle, wobei der Export eine Begrenzung auf jeweils 10.000 Treffer erzwingt. Im Sinne einer angemessenen qualitativen Analyse war daher eine Beschränkung des Korpusumfangs unumgänglich. Berücksichtigt wurden schließlich nur die Jahrgänge 2000 und 2022, also der älteste und der zum Zeitpunkt der Erhebung (März 2024) letzte zugängliche Jahrgang. Das bietet einerseits den Vorteil einer insgesamt beherrschbaren Datenmenge, andererseits können ansatzweise diachrone Entwicklungen im Zeitraum von zwei Jahrzehnten sichtbar gemacht werden. Der Jahrgang 2022 dürfte die aktuelle Praxis gendersensiblen Sprachgebrauchs vor dem Hintergrund eines mehrjährigen metasprachlichen Diskurses widerspiegeln. Das taz-Teilkorpus umfasst 17.113.331 (2000) bzw. 12.237.101 (2022) Tokens, insgesamt also 29.350.432 Wörter.

Die Menge von 20.000 extrahierten Treffern war zunächst manuell zu bereinigen. Aussortiert wurden auf *-ende(n)* auslautende Substantive (*Abende*, *Samenspende*,

⁶ <https://www2.ids-mannheim.de/cosmas2/web-app/hilfe/suchanfrage/eingabe-zeile/thematische-bsp/bsp-gendersprache.html> (aufgerufen am 23. Juni 2024).

Wochenenden), adjektivische Partizipien mit Großschreibung (*Verbrauchende Forschung*, *Vorsitzender Richter*), zu denen auch Kurzformen gehören (*Regierender* [Bürgermeister]). Dabei erwiesen sich im Einzelfall vermeintliche Adjektive als remotivierte Formen und damit als Substantive, die zu berücksichtigen waren (z. B. *Damit aus Nichtssagenden Neinsager werden*). Getilgt wurden auch Substantive, denen synchron kein Verb zuzuordnen ist (*Anwesende*) oder die als sprachspielerische Scherzbildungen anzusehen sind.⁷ Unberücksichtigt blieben schließlich substantivierte Partizipien, deren verbaler Charakter erhalten geblieben ist, da im Rahmen erweiterter Partizipialattribute Ergänzungen realisiert wurden (z. B. *die zunächst bei ihnen Schlafenden; die unter ihnen „Mitreisenden“; [...] der als Identifikationsfigur für alle ohnmächtig Wütenden wütet [...]*). Diese Konversionsprodukte sind nicht äquivalent mit deverbale Nomina agentis, weshalb sie auszuschließen waren.

3.2 Gegenwartssprachliche Daten: taz 2000 und 2022

Für das taz-Teilkorpus 2000 ergibt sich nach der Bereinigung ein Beleginventar von 4.236 Wortformen, die als konvertierte Partizipien I mit Personenreferenz zu klassifizieren sind. Diesen Belegen liegen 353 Types zugrunde, die in unterschiedlicher Frequenz gebraucht werden. Die Untersuchung der Types ist allerdings wenig aussagekräftig, da es sich in vielen Fällen um komplexe Bildungen und somit um Instanziierungen weniger Muster handelt; das gilt beispielsweise für die zahlreichen Komposita mit dem Zweitglied *Vorsitzende(r)*. Daher sind zunächst die Musterabstraktionen (Lemmata) zu ermitteln. Als Lemma wird das einzelne Konversionsprodukt zu Partizipien monomorphematischer Verben (*Forschende*, *Lesende*) und Präfix- sowie Partikelverben (*Vererbende*; *Einkaufende*) betrachtet. Bei komplexen Partizipien, die als Komposita zu analysieren sind, wird als Lemma dementsprechend das rechte Element angesetzt (*Fraktions-*, *Harems-*, *Landtagsfraktionsvorsitzende*); das gilt auch, wenn das dem Partizip zugrunde liegende Verb auf ein Syntagma zurückgeht: Univerbierungen wie *Grundwehrdienstleistende*, *Kleingewerbetreibende*, *Regieführende* usw. wurden als Instanziierungen eines Musters

⁷ Das betrifft beispielsweise *Führerscheinleistende* (taz 07.01.2000) und *Gewerdender* („Damit wird alles klar: Das Nichtsein ward im Gesein des Unhabens zum Abwesen als einem gewestem Gewerdenden – und vice versa“; taz 18.07.2000).

X-Leistende usw. betrachtet.⁸ Bei abweichender Semantik wurden homonyme Lemmata angesetzt, das gilt beispielsweise für ¹*Führende* („die in Führung liegen“) und ²*Führende* („die Aufsicht/Regie führen“).

Insgesamt lassen sich die 4.236 Belege auf 186 Lemmata in diesem Sinne verteilen. Dabei dominieren erwartungsgemäß einige wenige Muster, wie der Diversifikationsquotient von 0,044 zeigt. Allein auf die fünf Lemmata mit den meisten Realisierungen (*Vorsitzende*, *Studierende*, *Reisende*, *Überlebende*, *Suchende*) entfallen 3.271 Belege, das sind 77,2 % aller festgestellten Tokens.

Im Jahrgang 2022 sind 4.515 Tokens belegt, die sich auf 458 Types verteilen. Gegenüber dem Jahrgang 2000 ist somit eine quantitative Zunahme der Belege erkennbar, nicht nur hinsichtlich der absoluten Zahlen, sondern insbesondere in Relation zum jeweiligen Korpusumfang: Beträgt die relative Häufigkeit im Jahrgang 2000 noch 247,53 pMW, sind es 2022 bereits 368,96 pMW.

Zu dieser Entwicklung passt, dass sich auch die Vielfalt der Ausgangsformen erhöht hat. Die 4.515 Belege verteilen sich auf nunmehr 250 Lemmata; damit ergibt sich ein Diversifikationsquotient von 0,055. Zwar dominieren auch hier Varianten der Lemmata *Vorsitzende* und *Studierende*, ihr Anteil ist aber geschrumpft, und auf die fünf Lemmata mit den meisten Realisierungen (*Vorsitzende*, *Studierende*, *Überlebende*, *Mitarbeitende*, *Protestierende*) entfallen nur noch 2.320 Belege und damit 51,4 % aller festgestellten Tokens.

Diese meisten dieser hochfrequenten Lemmata sind hochgradig lexikalisiert und insofern unauffällig, als sie mit einer Ausnahme (*Mitarbeitende*) nicht mit deverbalen *-er*-Ableitungen konkurrieren; für *Studierende* existiert zudem eine Konkurrenzform *Studenten*.

Unproblematisch und für die Analyse irrelevant sind Konversionsprodukte geringer Frequenz, zu denen keine Nomina agentis belegt sind. Den Beispielen (22)–(24) ist gemeinsam, dass sie eine kontextuell gebundene Gleichzeitigkeitslesart evozieren:

⁸ Als Zitatform wird im metasprachlichen Text im Sinne der Einheitlichkeit jeweils die starke Pluralform (*-ende*) genutzt.

- (22) [...] oder Michael Rutschky, der beobachtet, wie Leute im vollen Mehrzweckbecken bumsen, und daraus sofort eine Theorie der Transformation öffentlicher Sitten bastelt – und der natürlich nicht ahnt, dass die **Bumsenden** ihren Terror der Intimität nur für ihn inszeniert haben. (taz 18.07.2000)
- (23) Der Duft der Hackepeterbrötchen stach mir in die Nase und lenkte von Mohrs Landschaftsaquarellen über den **Kauenden** ab. (taz 04.05.2000)
- (24) Sitzende, Sterbende, Streikende, Wartende, Winkende

Für die Fragestellung des Beitrags sind diejenigen konvertierten Partizipien relevant, für die prinzipiell auch ein deverbales Nomen agentis zur Verfügung steht.

Ob das der Fall ist, wurde systematisch anhand des Duden-Online-Wörterbuches geprüft; so konnten auch wenig frequente Lexeme ermittelt werden (*Zelter/Zelterin*). In den meisten Fällen handelt es sich um Derivate auf *-er*; bei einigen Stämmen kommen auch andere Ableitungsmuster in Betracht (*Habilitand, Akteur, Nachkomme*).

In einige Fällen kommt es zu semantischen Abweichungen, weil sich das Nomen agentis auf eine bestimmte, z. T. metaphorische Verbbedeutung bezieht (*Aussteiger, Einsteiger*), auf eine professionelle Lesart festgelegt ist (*Einkäufer, Erzieher*) oder durch ein homonymes Nomen instrumenti blockiert ist (*Fernseher*). Partizip und Nomen agentis können sich auch durch eine modale Bedeutungskomponente unterscheiden, wie die Belege (25) und (26) illustrieren.

- (25) a. Unter denen, die jetzt die Schwerhörigenschule besuchen, würden sicher viele davon profitieren, mit **Hörenden** unterrichtet zu werden. (taz 09.06.2000)
- b. #Unter denen, die jetzt die Schwerhörigenschule besuchen, würden sicher viele davon profitieren, mit **Hörern** unterrichtet zu werden.
- (26) a. Auch wenn man als **Hörende** nicht erfasst, wie sie eigentlich entstehen, sind sie auf dem Doppelalbum „Amaryllis/Belladonna“ vollkommen eingelöst. (taz 01.07.2022)
- b. Auch wenn man als **Hörerin** nicht erfasst, wie sie eigentlich entstehen, sind sie auf dem Doppelalbum „Amaryllis/Belladonna“ vollkommen eingelöst.

Eine echte Konkurrenz zwischen Partizipkonvertat und Nomen agentis besteht bei jeweils 54 (2000) bzw. 121 (2020) Lemmata; nur diese werden berücksichtigt. Für die beiden Jahrgänge im taz-Teilkorpus ergibt sich die Gesamtzahl von 138 Lemmata. In (27a) sind diejenigen aufgelistet, die primär mit stammgleichen *-er*-Derivaten konkurrieren; (27b) zeigt die Partizipien, bei deren Konkurrenzformen andere Ableitungsmuster relevant sind:

- (27) a. Anbietende; Angreifende; Anrufende; Antragsstellende; Anwohnende; Arbeitnehmende; Aufstockende; Ausbildende; Ausstellende; Begleitende; Beherrschende; Beitragende; Beobachtende; Beratende; Berichterstattende; Beschreibende; Beschwerdeführende; Besitzende; Besuchende; Betende; Betrachtende; Betreuende; Bewerbende; Bewohnende; Bietende; Blockierende; Darstellende; Einladende; Einwohnende; Empfangende; Erzählende; Fahrende; Falschparkende; Forschende; Fragende; Gastgebende; Gebende; Gestaltende; Golfende; Hassende; Haushaltende; Heimkehrende; Helfende; Herrschende; Hörende; Interviewende; Kämpfende; Kaufende; Klagende; Laufende; Lauschende; Leistungsbeziehende; Leitende; Lernende; Lesende; Leugnende; Machende; Machthabende; Mahnende; Meldende; Mietende; Mitarbeitende; Mitfahrende; Mithörende; Mitspielende; Mitstreitende; Nachfolgende; Nutzende; Nutznießende; Pfeifende; Pflegende; Querdenkende; Quereinsteigende; Rechtsabbiegende; Redende; Rufende; Saufende; Schauspielende; Schenkende; Schiedsrichtende; Schlafende; Schreibende; Schwärmende; Schwimmende; Seelsorgende; Skatende; Spielende; Spottende; Sprechende; Suchende; Betreibende; Tanzende; Teilnehmende; Telefonierende; Träumende; Turnende; Unterstützende; Unterzeichnende; Urlaubende; Veranstaltende; Verbrauchende; Verdienende; Vererbende; Verkaufende; Vermietende; Versorgende; Vertretende; Vorlesende; Vorstehende; Werbende; Zahlende; Zeltende; Zuhörende; Zuschauende; Zusehende
- b. Agierende; Assistierende; Asylbeantragende; Demonstrierende; Diskutierende; Dozierende; Erbende; Flüchtende; Habilitierende; Inserierende; Kandidierende; Kommandierende; Konsumierende; Kritisierende; Kuratierende; Moderierende; Neuabonnierende; Organisierende; Produzierende; Promovierende; Protestierende; Rebellierende; Studierende

Die folgenden exemplarischen qualitativen Analysen konzentrieren sich auf die Lemmata, für die im taz-Teilkorpus (a) mindestens zehn Instanziierungen belegt sind und die (b) mit einem deverbale Nomen agentis auf *-er* konkurrieren; Fremdwörter (z. B. *Demonstrant*) und Bildungen mit anderen Suffixen (z. B. *Flüchtling*) bleiben außer Betracht.

Wie oben (Abschnitt 2.1) dargelegt wurde, sollte eine angemessene Beschreibung der Partizipien nicht nur die Gleichzeitigkeitsrelation berücksichtigen, sondern auch die jeweils anzusetzende Prädikatsklasse. Daraus ergibt sich eine Differenzierung der Semantik der deverbale Personenbezeichnungen (vgl. Abschnitt 2.2). In Tab. 1 sind die fraglichen Partizipien, die die genannten Kriterien (a) und (b) erfüllen, in der Reihenfolge ihrer Belegfrequenz im taz-Teilkorpus dokumentiert. Ermittelt wurde in jedem Fall, ob eine Gleichzeitigkeitslesart in Betracht kommt, außerdem wurden die Prädikatsklasse und die semantische Charakterisierung bestimmt. Bei abweichender Klassifikation wurden

homonyme Lemmata angesetzt. Die Klassifikationen werden im Anschluss anhand ausgewählter Beispiele erläutert und kommentiert.

Tab. 1: Semantische Klassifikation der meistbelegten Partizipkonvertate in den untersuchten taz-Jahrgängen

Lemma	Frequenz	Gleichzeitigkeit	Prädikatsklasse	Bedeutung
<i>Mitarbeitende</i>	220	nein	EIGENSCHAFT	professionell
<i>Teilnehmende</i>	142	nein	EIGENSCHAFT	aktuell
<i>Forschende</i>	80	nein	HANDLUNG	professionell
<i>Verdienende</i>	76	nein	VORGANG	habituell
<i>Fahrende</i> ¹	63	nein	EIGENSCHAFT	habituell
<i>Fahrende</i> ²	3	ja	HANDLUNG	aktuell
<i>Herrschende</i>	43	nein	EIGENSCHAFT	professionell
<i>Pflegende</i>	26	nein	HANDLUNG	professionell
<i>Lesende</i> ¹	15	ja	HANDLUNG	aktuell
<i>Lesende</i> ²	10	nein	HANDLUNG	habituell
<i>Zuschauende</i>	21	ja	HANDLUNG	aktuell
<i>Betrachtende</i> ¹	17	ja	HANDLUNG	aktuell
<i>Betrachtende</i> ²	1	nein	HANDLUNG	aktuell
<i>Schreibende</i>	18	nein	HANDLUNG	professionell
<i>Anwohnende</i>	16	nein	EIGENSCHAFT	habituell
<i>Helfende</i>	15	ja	HANDLUNG	aktuell
<i>Besitzende</i>	14	nein	EIGENSCHAFT	habituell
<i>Zuhörende</i>	12	ja	HANDLUNG	aktuell
<i>Tanzende</i>	11	ja	HANDLUNG	aktuell, habituell
<i>Nutzende</i>	10	nein	EIGENSCHAFT	habituell

Wie Tab. 1 zeigt, ist nur bei einem kleinen Teil der belegten Partizipien eine Gleichzeitigkeitslesart anzusetzen. In den meisten Fällen dienen die fraglichen Bildungen dem Ausdruck der Zugehörigkeit zu einer Personengruppe, sie markieren einen STATUS, weshalb ihnen die Prädikatsklasse EIGENSCHAFT zukommt; das gilt beispielsweise für *Mitarbeitende*, *Teilnehmende*, *Pflegende* und *Schreibende*. So ist *Mitarbeitende* äquivalent mit *Beschäftigte* (28); in Beleg (29) ist die Gleichzeitigkeitsrelation sogar logisch ausgeschlossen.

- (28) Medienhäuser, die sich dem Kodex angeschlossen haben, erklären sich dazu bereit, bestimmte Standards zum Schutz ihrer **Mitarbeitenden** umzusetzen. (taz 27.01.22)

- (29) Neulich habe ich über ein Bürgeramt in Mitte geschrieben, in dem ich elektronisch einen Termin für die Beantragung eines neuen Reisepasses gebucht hatte, dann aber wegen der offenbar gleichzeitigen Mittagspause aller **Mitarbeitenden** fast 45 Minuten hatte warten müssen. (taz 26.07.22)

Der Charakter einer Statusbezeichnung spiegelt sich auch einer ganzen Reihe entsprechender Komposita wider:

- (30) Bundestags-, Fakultäts-, Feuerwehr-, Flughafen-, Justiz-, Küchen-, Projekt-, Rettungsdienstmitarbeitende

Gleiches gilt für Partizipien, die auf HANDLUNGEN im Sinne regelmäßig wiederholter Tätigkeiten referieren:

- (31) Die **Pflegenden** kommen gar nicht mehr dazu, sich einfach um Menschen mal zu kümmern, mit denen mal in den Park oder Einkaufen zu gehen. (taz 19.05.22)
- (32) Die **Forschenden** sehen Erfolge mit einzelnen Patient*innen und sind deshalb nicht mehr neutral. (taz 23.07.22)
- (33) Das Tunguska-Ereignis wirft bis heute Rätsel auf, viele **Forschende** gehen jedoch davon aus, dass es sich um einen Asteroiden handelte, der am Morgen des 30. Juni 1908 durch die Erdatmosphäre brach. (taz 06.08.22)
- (34) Hunger wird in der schönggeistigen Literatur kaum behandelt. Vielleicht weil Hungernde selten schreiben und **Schreibende** selten hungern. (taz 22.06.22)
- (35) Nun hat eine Macht immer schon mitbestimmt über Wohl und Wehe der **Schreibenden**: der Markt. (taz 09.03.22)

Die Bedeutung lässt sich wie die der äquivalenten Nomina agentis als ‚professionell‘ kennzeichnen; es geht nicht darum, dass jemand aktuell ‚am Pflegen/Forschen/Schreiben ist‘.

Eine habituelle Semantik ist bei Partizipien anzusetzen, die Personen als Träger von Eigenschaften bezeichnen, das gilt etwa für *Anwohnende* (36–37):

- (36) Gegen die Umbenennung einer Straße können sich nur **Anwohnende** juristisch wehren. (taz 19.08.22)
- (37) Sie ist Geoökologin, **Anwohnende** und Gärtnerin. (taz 14.06.22)

Zu den Eigenschaftsprädikaten zählt auch *nutzen* in der Bedeutung ‚Kunde sein‘; das abgeleitete Partizip verweist auf diesen zeitlich nicht näher eingegrenzten Status.

- (38) Zunächst führten sie eine Online-Erhebung mit 12.000 Teilnehmenden durch, um das Verhalten von **Nutzenden** und Nichtnutzenden vergleichen zu können. (taz 15.01.22)
- (39) Die Plattformen, die seit Jahren **Nutzende** gegeneinander aufhetzen, sollen nicht etwa stärkere Pflichten gegenüber den sie Nutzenden erhalten, sondern zu Handlangern der Politik gemacht und zum Verrat der sie Nutzenden verpflichtet werden. (taz 28.05.22)
- (40) Aber Leipzig ist berühmt als fahrradfreundliche Stadt und der MDR für seine Nähe zur Lebenswirklichkeit seiner **Nutzenden**. (taz 02.06.22)

Es sind aber auch Partizipien belegt, die durchaus eine Gleichzeitigkeitslesart nahelegen. Das betrifft von Handlungsverben abgeleitete Bildungen wie *Helfende* (41) und *Zuhörende* (42–43), wobei sich für (42) strenggenommen ein Paradoxon ergibt (worauf wohl in der Parenthese angespielt wird).

- (41) Zwischen **Helfenden** in Warnwesten, Sicherheitsleuten, Feuerwehrmännern und Polizisten, [!] rennen Kinder herum. (taz 18.05.22)
- (42) Die Scheinwerfer verlassen die Bühne und man sieht – als eigene Inszenierung – die **Zuhörenden** sich die Ohren zuhalten. (taz 23.05.00)
- (43) Und die **Zuhörenden** verändern gemäß ihren Möglichkeiten den emotionalen Sinn des Gehörten, um verstehen zu können, und glauben dann, verstanden zu haben. (taz 24.05.00)

Diese Partizipien könnten aber ohne Bedeutungsverlust durch Nomina agentis wie *Helfer* oder *Zuhörerinnen* ausgetauscht werden, weil auch diese Bildungen eine Lesart im Sinne von ‚Person, die gerade hilft/zuhört‘ erlauben. Der Ausdruck der Gleichzeitigkeitsrelation ist demnach kein exklusives semantisches Merkmal der substantivierten Partizipien.

Dass die Bedeutung des Partizips sich nicht aus den formalen Merkmalen des Bildungsmusters ergibt, sondern von der Semantik des Verbs abhängt, lässt sich gut an den folgenden Beispielen feststellen. Dem Partizip *lesend* – und damit auch seiner Substantivierung *Lesende* – kann man mit Eisenberg (2020a: 269) in aspektueller Hinsicht grundsätzlich eine imperfektive Lesart zuordnen und Gleichzeitigkeit mit dem übergeordneten Geschehen unterstellen. Das illustrieren die Belege (44)–(46), hier geht es um Personen, die sich aktuell im Leseprozess befinden:

- (44) Aber, und das ist die große Kunst von Khider, „Der Erinnerungsfälscher“ spielt nicht nur mit Wahrheit und Erzählung, Dichtung und Realität, sondern ebenso geschickt mit den Erwartungen des **Lesenden**, der sich gerade denkt: Eigentlich ist das doch ein Märchen. (taz 01.03.22)
- (45) Die Lebendigkeit der Figuren entsteht durch Thomas' Nähe zu ihnen, in die sie die **Lesenden** von Beginn an hineinzuziehen vermag. (taz 16.04.22)
- (46) Eine kleine Energiesparleuchte drückt sich ins angeheftete Kissen und strahlt, vom weißen Stoff reflektiert, für den **Lesenden**. (taz 01.04.00)

Die Partizipien sind hier besonders geeignet, die Gleichzeitigkeit der Lesehandlung explizit zu markieren. Der Gebrauch eines -er-Derivats wäre allerdings in allen drei Fällen problemlos und ohne Bedeutungsverlust möglich. Nun ist *lesen* aber nicht nur als Handlungsverb zu klassifizieren; Verben lassen sich verschiedenen Prädikatsklassen zuordnen, wie v. Polenz (2008: 166) am Beispiel von *sehen* illustriert. Für *lesen* gilt, dass es auch als Eigenschaftsverb fungiert (‚regelmäßig/gern/nicht lesen‘), und diese Bedeutung liegt dem Partizip in (47) zugrunde:

- (47) Denn so wie die große taz-Community aus **Lesenden**, Genoss*innen, Abonnent*innen und anderen Unterstützer*innen vom intensiven Austausch lebt, braucht auch die Community der taz-Mitarbeitenden den kreativen Austausch [...]. (taz 28.05.22)

Die *Lesenden* sind hier diejenigen, die die betreffende Zeitung regelmäßig rezipieren, nicht unbedingt zum Äußerungszeitpunkt; der Ausdruck verweist also – ebenso wie *Mitarbeitende* – auf die Zugehörigkeit zu einer Gruppe. Diese Polysemie kennzeichnet auch das Nomen agentis *Leser*. Bemerkenswert ist hier, dass die übrigen Personenbezeichnungen, die in der Präpositionalphrase koordiniert sind, dem Bildungsmuster *X*innen* folgen; ob hier zwischen *Lesenden* und *Leser*innen* differenziert werden soll oder nur Variation (vgl. Abschnitt 3.3) beabsichtigt ist, lässt sich nicht feststellen.

Vergleichbar sind die Verhältnisse bei einem anderen Verbum sentiendi: *Betrachtende* sind diejenigen, die ein Objekt zu einem bestimmten Zeitpunkt genau ansehen bzw. auf sich wirken lassen; fokussiert wird die Zeitspanne der Wahrnehmung (48–50).

- (48) Wer einmal vor den riesigen, schwarzen Ölbildern des 101 Jahre alten, französischen Malers stand, wird Flecks Argument nachvollziehen können, dass die Reflexe, die von der Oberfläche dieser Werke ausgingen, so unmittelbar das Nervensystem der **Betrachtenden** erreichten wie digitale Effekte. (taz 11.01.22)

- (49) Ein älteres Foto zeigt sie mit langen Haaren, Rollkragenpullover und dem **Betrachtenden** zugewandt. (taz 05.07.22)
- (50) Lili Fischers Gesichtsmuskeln geben einiges her und eröffnen so dem **Betrachtenden** im wahrsten Sinne des Wortes tiefe Einblicke. (taz 16.02.00)

Beleg (51) weicht davon ab, es liegt die Bedeutung ‚zum Gegenstand einer genauen Untersuchung, Beurteilung machen‘ (Duden o. J. s.v. *betrachten*) zugrunde, eine strenge Gleichzeitigkeitsrelation liegt nicht vor. Auffällig ist zudem die syntaktische Einbindung; das substantivierte Partizip erhält ein Attribut in Form eines Genitivus obiectivus, was einen gewissen Lexikalisierungsgrad voraussetzt.

- (51) Der arg eurozentrische Blick auf den Fußball schleicht sich auch beim **Betrachtenden** der Frauen-EM schnell ein. (taz 21.07.22)

Der Typ *Fahrende* erlaubt grundsätzlich eine Gleichzeitigkeitslesart, wie Beleg (52) zeigt; die betreffenden Personen vollziehen zu einem Referenzzeitpunkt eine Handlung:

- (52) Der Abstand beim Überholen von **Fahrradfahrenden** durch Kraftfahrzeuge ist geringer, wenn sie einen Helm tragen. (taz 16.07.22)

Diese Verwendungsweise ist jedoch die Ausnahme. In (53) scheint Gleichzeitigkeit sogar logisch ausgeschlossen, weil man nicht zugleich fahren und jemanden von der Straße schubsen/ziehen kann:

- (53) Wütende **Autofahrende** schubsten und zogen die Aktivist*innen von der Straße. (taz 15.07.22)

Die Konstellation dieses Belegs ähnelt auf den ersten Blick den in der Sprachkritik diskutierten konstruierten Beispielen wie *schlafende Dozierende* usw. (vgl. Payr 2022: 97), sofern eine Gleichzeitigkeitslesart unterstellt wird. Allerdings besitzt die fragliche Tätigkeit eine zeitliche Ausdehnung und kann angemessener als „Kette kombinierter Teilhandlungen“ (v. Polenz 2008: 160) beschrieben werden (‚einen Weg mit dem Auto zurücklegen‘; ‚mit dem Auto unterwegs sein‘). Zumeist ist aber eine habituelle Lesart intendiert; die Partizipien dieses Typs dienen in der Regel der Referenz auf eine Personengruppe (54). In diesen Fällen ist von der Prädikatsklasse EIGENSCHAFT auszugehen.

- (54) Leider kam der Zuwachs an **Radfahrenden** stärker aus dem Lager der ÖPNV-Nutzer*innen als aus dem Lager der **Autofahrenden**. (taz 14.03.22)

Ähnlich uneinheitlich ist die Semantik des Partizips *Tanzende* zu beschreiben. Im Sinne der Auffassung Eisenbergs wird es in (55) gebraucht; es ist von Personen die Rede, die zum Referenzzeitpunkt am Tanzen sind:

- (55) Zur Zugabe kamen sie nämlich von der Bühne herunter zwischen die **Tanzenden** [...] und spielten dort schwitzend im Gewimmel fast noch einen kompletten dritten Set [...]. (taz 10.01.00)

Dieser Gebrauch hat eine Entsprechung in (56); bemerkenswert ist hier, dass im unmittelbaren Kontext zusätzlich auch das mit Binnen-I genderte Nomen *agentis TänzerInnen* gebraucht wird.

- (56) Die Choreographien waren immer auf die Überraschung ausgerichtet, man konnte nie vorhersehen, welche (körperliche und dramaturgische) Drehung die TänzerInnen als nächstes ausführen würden. Dabei war, zumindest beim ersten Stück, die Lichtregie genauso virtuos wie die **Tanzenden**. (taz 25.04.00)

Es ist möglich, dass hier semantisch differenziert werden soll und bei *TänzerInnen* die Lesart ‚professionell‘ im Vordergrund steht, während *Tanzende* auf die aktuelle Tätigkeit verweist. Zwingend ist diese Interpretation jedoch nicht; die Variation kann auch dem Bemühen geschuldet sein, Wiederholungen zu vermeiden.

Wie schon in anderen Fällen zeigt sich auch für *Tanzende*, dass der Ausdruck auch dann gebraucht wird, wenn eine Gleichzeitigkeitsinterpretation nicht in Frage kommt (57).

- (57) Zur Abkühlung der **Tanzenden** wurde an einer Stelle sogar ein Gartenschlauch in einer Grünanlage angeschlossen, was auch von den anwesenden Eltern zur Bespaßung ihrer Kinder dankend angenommen wurde. (taz 25.07.22)

Ein semantischer Unterschied zum -er-Derivat besteht jenseits der Genusneutralisierung nicht; das Partizip leistet das Gleiche wie eine kontextgebundene deiktische Bildung *Tänzer(innen)* (‚die gerade getanzt haben‘).

Aus den Befunden ergibt sich folgendes Bild:

- Die belegten substantivierten Partizipien lassen sich auf Verben unterschiedlicher Prädikatsklassen zurückführen.
- Die (intendierte) Semantik ist nicht einheitlich.

- Eine Gleichzeitigkeitslesart ist in einigen Fällen möglich, sie ist aber nicht zwingend mit dem Bildungsmuster verbunden. Eine 1:1-Entsprechung von Form und Bedeutung ist nicht erkennbar.
- Diese Vielfalt gilt auch für deverbale Nomina agentis, deren Bedeutung ebenfalls kontextgebunden ist.
- Die Partizipien sind jeweils mit *-er*-Derivaten (bzw. den entsprechenden movierten Formen) äquivalent; ein Austausch wäre ohne Bedeutungsverlust möglich.

Die untersuchten Belege legen nahe, dass die substantivierten Partizipien als geschlechtsneutrale Varianten der Pluralformen deverbaler Nomina agentis gebraucht werden. Ihre Bildung ist im Vergleich zu maskulinen *-er*-Ableitungen mit einem höheren Aufwand verbunden (gemessen an der Silbenzahl); bei der taz ist aber in Rechnung zu stellen, dass der Gebrauch genderinklusive Formen ohnehin einen höheren Stellenwert einnimmt. Andererseits haben die Partizipien kostenseitig sogar einen Vorteil, wenn man ihnen andere gegenderte Ausdrücke wie Beidnennungen oder Pluralformen der Derivate auf *-*in* gegenüberstellt. Gleichwohl ist ein Vergleich unter sprachökonomischen Aspekten wenig zielführend, weil dann pragmatische Aspekte der „Signalkommunikation“ (Feilke 2023: 1) und Angemessenheit ebenso unberücksichtigt blieben wie die mit den einzelnen Formen verbundenen Intentionen (vgl. Kotthoff 2020; Müller-Spitzer 2022).

Als Hypermaxime sprachlichen Handelns im Sinne Kellers (2014: 142) kann vielmehr gelten: „Rede so, daß Du sozial erfolgreich bist.“ Keller (2014: 142) merkt dazu an: „Was jeweils als erfolgreich angesehen wird, ist von Fall zu Fall verschieden, von Situation zu Situation, von Individuum zu Individuum, von Gruppe zu Gruppe, von Adressat zu Adressat.“ Das Ökonomieprinzip greift diesbezüglich also nicht.

Man könnte aber weitere Vorteile geltend machen – jeweils unter der Annahme, dass die Partizipien nicht per se einen semantischen Mehrwert haben bzw. eine Gleichzeitigkeitslesart erzwingen:

- Der Bildungstyp Konversion unterliegt praktisch keinerlei Beschränkungen, während die Bildung von *-er*-Derivaten aus unterschiedlichen Gründen blockiert sein kann (vgl. Wurzel 1988).
- Durch den morphologischen Marker *-end-* ergibt sich ein einheitliches Kennzeichen für Substantive derselben Bedeutung. Auch Fremdwörter können so demselben Bildungsmuster unterzogen werden (*Demonstrierende*), was die Reihenbildung erleichtert und die morphologische Integration befördert (*Studierende, Lehrende, Promovierende*). Damit entsteht größere Eindeutigkeit – einerseits in morphologischer

Hinsicht, zumal beim konkurrierenden *-er*-Suffix größere Vielfalt hinsichtlich der Bedeutung herrscht (vgl. Scherer 2005: 91), andererseits hinsichtlich der Intention, da die betreffenden Formen Signalwirkung besitzen (vgl. Harnisch 2016, Feilke 2023).

Zur angemessenen Interpretation der Befunde ist es aber erforderlich, die Ergebnisse in einen größeren Zusammenhang zu stellen. Daher soll im nächsten Schritt die Eignung der Partizipien zur genderneutralen Personenreferenz geprüft werden; dazu sind auch Konkurrenzformen in den Blick zu nehmen (vgl. auch Müller-Spitzer et al. 2024).

3.3 Partizipien im Vergleich mit anderen genderten Formen

Wie bereits dargestellt, bieten substantivierte Partizipien die Möglichkeit, genusneutrale Personenbezeichnungen zu realisieren – vorausgesetzt, sie stehen im Plural (vgl. Kotthoff & Nübling 2024: 175). Bei den im Teilkorpus taz 2022 belegten Partizipien handelt es sich überwiegend um Pluralformen; nur 1.112 Belege (25 %) stehen im Singular. Das ist insbesondere bei spezifischer Referenz der Fall und betrifft hauptsächlich Rollen-/Funktionsbezeichnungen von Individuen; allein beim hochfrequenten Typ (X-) *Vorsitzende* sind es 977 Fälle. Das Genus hängt dann vom Geschlecht der Person ab, auf die jeweils referiert wird. Wie die folgenden Beispiele⁹ zeigen, werden Singularformen bei spezifischer (58), aber auch bei generischer bzw. distributiver Referenz (vgl. Eisenberg 2020b: 170–171, 175) gebraucht; die Genusneutralität ist – anders als bei Pluralformen – nicht mehr gegeben.

- (58) Nicht jede*r, der die 112 anruft, bekommt automatisch einen Rettungswagen geschickt. Manchmal entscheiden sie in der Leitstelle auch, dass **der Anrufende** kein Notfall ist. (taz 12.11.2022)
- (59) Ein Raunen ging gestern durch die Fachabteilung, als es auch **dem letzten Baumarkt-besuchenden** klar wurde, dass es gar nicht „Imbusschlüssel“ heißt. (taz 19.01.22)
- (60) Der arg eurozentrische Blick auf den Fußball schleicht sich auch **beim Betrachtenden** der Frauen-EM schnell ein. (taz 21.07.22)
- (61) Es scheint bei **jedem Flughafenmitarbeitenden** eine Gruppe Reisende zu stehen, die wild mit ihren Tickets gestikulieren. (taz 15.07.22)
- (62) So wird die bittere, menschenfeindliche Realität der Flucht erträglich, ja nahezu weichgezeichnet – für den Geflüchteten selbst, aber auch für **den Lesenden**. (taz 01.03.22)

⁹ Die Liste ist keineswegs vollständig; vgl. auch die bereits diskutierten Belege (44), (46), (49), (50).

- (63) Dieser wolle durch die Nummerierung sicherstellen, dass **jeder Mietende** nur einen Antwortfragebogen übermittle – zudem würden sämtliche personenbezogenen Daten nach dem Auftrag gelöscht, versichert der Konzern. (taz 10.05.22)

Gebraucht wird jeweils die maskuline Form im Singular, die wohl geschlechtsübergreifend zu verstehen ist. Hier scheint sich Harnischs (2016: 169) Vermutung zu bestätigen, dass die Wahl des Partizips anstelle eines *-er*-Derivats zunächst eine pragmatische Funktion besitzt, nämlich eine Präferenz für (potenziell) genderneutrale Formen anzuzeigen, ein Anliegen, das durch das generische Maskulinum letztlich konterkariert wird. Eine solche Inkonsistenz zeigt sich auch in Beleg (58), wenn zwar das Indefinitpronomen im Suffix ein <*> erhält, das folgende Relativpronomen aber als Maskulinum ausgewiesen ist. Hier deuten sich die textgrammatischen Probleme an, die sich bei Verwendung gegenderter Singularformen ergeben. In diesem Zusammenhang ist ein Blick auf andere Verfahren der Personenreferenz geboten, die in der taz genutzt werden. Insgesamt ist ein Bemühen um Variation (im Sinne von Feilke 2023) erkennbar, etwa wenn (lexikalisiertes) Partizip, Form mit Genderstern und generisches Maskulinum in einer Phrase koordiniert werden, wie es in (64) der Fall ist:

- (64) Tausende Lehrer*innen, Studierende und empörte Bürger sind am Sonntag gegen die Bildungspolitik der nationalkonservativen Regierung von Viktor Orbán marschiert. (taz 25.10.22)

Variiert wird auch in den folgenden Beispielen, wenn statt des Gendersterns ein <:> gebraucht (65) oder der Stern mit Binnen-<I> kombiniert wird (66).

- (65) Als erneut eine Abschiebung drohte, wandten sich auch 15 Professor:innen und Mitstudierende an Bundes- und Landespolitiker und forderten ein Bleiberecht für Mheddin Saho. (taz 21.02.22)

- (66) Wenn es nach den Gärtner*Innen und Anwohnenden geht, bleibt das auch so. (taz 14.06.22)

Ein klares System ist – auch angesichts der geringen Belegmenge – nicht erkennbar; es scheint aber, als werde das generische Maskulinum toleriert, sofern im umliegenden Text auch andere Formen genutzt werden (vgl. Kotthoff & Nübling 2024: 142 zu Streumaskulina); auch das kann als Beleg für die oben angesprochene pragmatische Funktion angesehen werden, als Ausdruck einer „Registrierung“ im Sinne von Kotthoff (2020).

Zur Einordnung der Gebrauchsfrequenz soll anhand ausgewählter Lemmata stichprobenhaft geprüft werden, welche Bedeutung den Partizipkonvertaten im Vergleich mit anderen Referenzstrategien überhaupt zukommt. Ausgewählt wurden die beiden Partizipien mit der gemäß Tab. 1 höchsten Frequenz im Korpus (*Mitarbeitende*, *Teilnehmende*) sowie mit *Lesende* und *Tanzende* exemplarisch zwei weniger frequente Lemmata, für die unterschiedliche Bedeutungsklassen anzusetzen sind und die z. T. Gegenstand der Kritik sind (vgl. Eisenberg 2021). Tab. 2 zeigt die Anzahl dieser im Jahrgang 2022 belegten Partizipien im Vergleich mit konkurrierenden Formen, nämlich den Bildungen auf *-*in* bzw. *-.in* (sowie ggf. *._in*), die hier als graphische Varianten¹⁰ zusammengefasst werden, der älteren Form mit Binnen-*I* sowie der Beidnennung. Erfasst wurden im Sinne der Fragestellung des Beitrags nur Pluralformen, da sie genusneutral sind; generische Maskulina bzw. Feminina blieben unberücksichtigt. Die Beschränkung auf den Jahrgang 2022 ist sinnvoll, weil er die gegenwärtige Praxis vor dem Hintergrund jüngerer metasprachlicher Diskurse widerspiegelt und auch die in jüngerer Zeit etablierten Formen mit Genderzeichen enthält. Zu beachten ist, dass die Zahl der Belege für die jeweiligen Partizipien in dieser Stichprobe zwangsläufig höher ausfällt als in der oben präsentierten Untersuchung: Da jetzt gezielt nach konkreten Lexemen gesucht werden konnte, konnten alle Treffer berücksichtigt werden; eine Reduktion des Gesamtmaterials auf 10.000 Treffer war nicht erforderlich. Ermittelt werden soll, in welchem quantitativen Verhältnis substantiviertes Partizip und Ausdrucksalternativen zueinander stehen.

¹⁰ Vgl. Völkening (2022) zum morpho-phonologischen Status der betr. Einheit sowie Ferstl & Nübling (2024: 261–273) zu einer Analyse der verwendeten Zeichen.

Tab. 2: Anteil verschiedener Bildungstypen für genderinklusive Referenzausdrücke im Vergleich (Jahrgang 2022)

Form	Anzahl Belege		Anteil (%)
<i>Mitarbeitende</i>		305	31,8
<i>Formen mit Genderzeichen insges.</i>		565	58,9
<i>Mitarbeiter*innen</i>	288		
<i>Mitarbeiter:innen</i>	275		
<i>Mitarbeiter_innen</i>	2		
<i>MitarbeiterInnen</i>		63	6,6
<i>Splittingformen insges.</i>		27	2,8
<i>Mitarbeiterinnen und Mitarbeiter</i>	25		
<i>Mitarbeiter und Mitarbeiterinnen</i>	2		
		960	100
<i>Teilnehmende</i>		174	34,1
<i>Formen mit Genderzeichen insges.</i>		308	60,4
<i>Teilnehmer*innen</i>	140		
<i>Teilnehmer:innen</i>	168		
<i>TeilnehmerInnen</i>		18	3,5
<i>Splittingformen insges.</i>		10	2,0
<i>Teilnehmerinnen und Teilnehmer</i>	8		
<i>Teilnehmer und Teilnehmerinnen</i>	2		
		510	100
<i>Lesende</i>		27	11,5
<i>Formen mit Genderzeichen insges.</i>		155	66,2
<i>Leser*innen</i>	92		
<i>Leser:innen</i>	62		
<i>Leser_innen</i>	1		
<i>LeserInnen</i>		29	12,4
<i>Splittingformen insges.</i>		23	9,8
<i>Leserinnen und Leser</i>	22		
<i>Leser und Leserinnen</i>	1		
		234	100
<i>Tanzende</i>		6	8,8
<i>Formen mit Genderzeichen insges.</i>		46	67,6
<i>Tänzer*innen</i>	19		
<i>Tänzer:innen</i>	27		
<i>TänzerInnen</i>		13	19,1
<i>Splittingformen insges.</i>		3	4,4
<i>Tänzerinnen und Tänzer</i>	2		
<i>Tänzer und Tänzerinnen</i>	1		
		68	100

Die Daten sind recht eindrucksvoll: Bei gegenderten Personenbezeichnungen dominieren die Bildungen auf *-*in* bzw. *-.in* (die sich nur graphisch unterscheiden und in morphologischer bzw. pragmatischer Hinsicht zusammengefasst werden können). Es gibt bei der taz also eine klare Präferenz für die Formen, die mit Kotthoff (2020: 114–120) dem „nicht binären Register“ zuzuordnen sind; nur Formen mit Unterstrich sind ausgesprochen randständig (vgl. auch Sökefeld 2021: 138). Substantivierte Partizipien machen im Fall der höherfrequenten Lexeme *Mitarbeitende* und *Teilnehmende* etwa ein Drittel der Belege aus; je seltener das Lemma, desto geringer ist der Anteil der Partizipbildungen. Daraus lässt sich schließen, dass in der taz die genannten Lemmata als weitgehend lexikalisiert gelten können; es sind zudem diejenigen, die in Tab. 1 als Eigenschaftsprädikate klassifiziert wurden. Auf der anderen Seite zeigt sich, dass Partizipien zu Handlungsverben wie *Lesende* und *Tanzende*, bei denen grundsätzlich auch eine Gleichzeitigkeitslesart in Frage kommt, insgesamt eher die Ausnahme darstellen.

3.4 Historische Daten: Historische Zeitungen und Zeitschriften (KHZ/KHM) 1737–1905

Aus dem historischen Teilkorpus wurde im ersten Schritt ein Beleginventar von 1.368 einschlägigen Wortformen extrahiert, die sich auf 233 Lexeme verteilen. Auch hier wurden nach den oben skizzierten Kriterien die zugrunde liegenden Lemmata ermittelt: Es ergeben sich 188 Lemmata. Das Verhältnis von Types zu Tokens (Diversifikationsquotient) beträgt 0,137; damit zeichnet sich eine größere Vielfalt an Bildungen als in den beiden taz-Jahrgängen ab. Das wird bestätigt durch den Blick auf die fünf Lemmata mit den meisten Realisierungen; in diesem Teilkorpus sind es *Reisende*, *Studierende*, *Vorsitzende*, *Treibende*, *Besitzende*. Auf sie entfallen 715 und damit nur 52,3 % der Tokens. Bemerkenswert ist gleichwohl, dass die drei häufigsten Lemmata – *Reisende*, *Studierende*, *Vorsitzende* – in beiden Teilkorpora dieselben sind. Es zeigt sich auch, dass *Studierende* bereits im 18. und 19. Jahrhundert eine etablierte Personenbezeichnung war. Bemerkenswert ist die relative Häufigkeit der Partizipkonvertate; sie beträgt 334,18 pMW – dieser Wert liegt über dem Mittelwert für die beiden taz-Jahrgänge (308,25). Bereits die quantitativen Daten zeigen, dass der Gebrauch substantivierter Partizipien zur Personenreferenz kein exklusives Merkmal der Gegenwartssprache ist und unabhängig vom Diskurs über genderinklusive Personenbezeichnungen zu betrachten ist.

Auch für das historische Teilkorpus wurden diejenigen Lexeme ermittelt, zu denen stammgleiche deverbale Nomina agentis bezeugt sind, die für den fraglichen Zeitraum als lexikalisiert gelten können. Maßstab dafür sind Belege in historischen Wörterbüchern, namentlich im *Deutschen Wörterbuch* der Brüder Grimm (DWB bzw. DWBN); ergänzend wurde das Wörterbuch von Adelung (1793–1801) herangezogen, für Fremdwörter das *Deutsche Fremdwörterbuch* von Schulz/Basler (DFWB). Bemerkenswert ist, dass einige der heute etablierten Nomina agentis in den einschlägigen historischen Wörterbüchern nicht verzeichnet sind; das ist ein Hinweis darauf, dass sie als Mittel zur Personenreferenz (noch) nicht allgemein etabliert waren und die betreffenden Partizipien den unmarkierten Fall darstellten. Das betrifft u. a. *Besuchende*, *Einzahlende*, *Verbrauchende*. Andererseits sind Belege für heute unübliche Nomina gebucht, beispielsweise *Vorsitzer* (DWB XII/2: 1589).

Einschränkend ist festzustellen, dass die Auskünfte nicht immer auf authentischen Sprachgebrauch schließen lassen, sondern u. U. die Buchungsprinzipien und idiosynkratische Vorstellungen des Lexikographen widerspiegeln können (vgl. Schulz 2000).¹¹ Als lexikalisiert wurden daher nur diejenigen Lexeme gewertet, zu denen auch authentische Textbelege genannt werden. Außer Betracht bleiben auch solche Nomina agentis, die ausdrücklich als ‚veraltet‘, ‚selten‘ oder ‚ungewöhnlich‘ ausgewiesen werden, beispielsweise *Versteher*; das DWB merkt hier an, es sei ein „ungewöhnliches subst. zu *verstehen* I, das erst in nhd. sprache erscheint. sonst wird der begriff durch das part. präs. von *verstehen* oder einen relativsatz umschrieben“ (DWB XII/1: 1701).

Nach den genannten Kriterien können die in (67) aufgelisteten 66 Lemmata von Partizipkonvertaten als äquivalent mit zeitgenössisch belegten Nomina agentis angesehen werden. Durch Fettdruck hervorgehoben sind die 20 Lemmata, die auch im synchronen taz-Korpus als äquivalente Varianten belegt sind.

¹¹ Dass Stieler, auf den im DWB immer wieder verwiesen wird, Lemmata wie *Sterber* (inkl. der movierten Form *Sterberinn*!) verzeichnet (1691: 2171), dürfte in erster Linie seinem Bemühen geschuldet sein, das Wortbildungspotenzial des Deutschen zu illustrieren und derartige Bildungen zu propagieren. Auf der anderen Seite stehen Lexeme wie *Verlierer*, *Verliererinn* (Stieler 1691: 1109), die sich haben durchsetzen können.

- (67) Anfragende; Annehmende; Anschauende; Ansuchende; **Arbeitgebende**; Ausführende; **Besitzende**; **Bietende**; **Commandirende**; Concurrirende; Darleihende; Denkende; Dienende; Erfüllende; Essende; **Fahrende**; **Fragende**; Gebärende; Gebietende; Gehende; Genießende; Harrende; Hellsehende; **Herrschende**; Inspicirende; **Kämpfende**; Kennende; **Klagende**; Lehrende; **Lesende**; **Mitarbeitende**; **Mitspielende**; Musizierende; Nähende; Obsiegende; Regierende; Reitende; Revidirende; Schauende; Schiffende; **Schlafende**; **Schreibende**; Spielende; **Sprechende**; Springende; Steuerzahlende; Strebende; **Studierende**; Stürmende; **Suchende**; **Tanzende**; **Theilnehmende**; Tragende; **Träumende**; Uebertreibende; Umwohnende; Urtheilende; Veranlassende; Vorsitzende; Wählende; Walzende; Widersprechende; Widerstrebende; Wissende; Zischende; **Zuschauende**

Über die 20 parallel belegten Partizipien hinaus sind weitere Lemmata in der Gegenwartsprache geläufig, beispielsweise *Reitende*, *Spielende*, *(Steuer-)Zahlende*, *Wählende*. Damit entsteht der Eindruck, diese Art der Personenreferenz sei im 19. Jh. bereits etabliert, und in der Tat finden sich zu dieser Zeit unerwartet ‚progressiv‘ anmutende Belege wie *Arbeitgebende* (68), das eine direkte Parallele im taz-Teilkorpus 2022 besitzt (69).

- (68) Werden die Einwanderer von ihren **Arbeitgebenden** selbst beköstigt, so erhalten sie natürlich auch frisches Fleisch. (Allgemeine Auswanderungs=Zeitung 16.03.1847)
- (69) So sollen die **Arbeitgebenden** Missstände abstellen können, ohne gleich am öffentlichen Pranger zu stehen. (taz 28.07.22)

Nun wird man allerdings nicht annehmen können, dass das Motiv die Vermeidung des generischen Maskulinums gewesen sei.¹² Das Bild relativiert sich auch, wenn man die absolute Anzahl der Belege mit der Gebrauchsfrequenz der stammgleichen Nomina agentis (die z. T. in movierter Form vorliegen) vergleicht; die entsprechenden Daten gehen aus Tab. 3 hervor:

Tab. 3: Substantivierte Partizipien im historischen Teilkorpus und ihre Entsprechungen

Partizipkonvertat	Beleganzahl	Nomen agentis	Beleganzahl
<i>Studi(e)rende</i>	113	<i>Student</i>	86
<i>Besitzende</i>	32	<i>Besitzer(in)</i>	231
<i>(Meist-)Bietende</i>	28	<i>x-Bieter</i>	0
<i>Commandirende</i>	19	<i>Commandant, -eur</i>	160

¹² Immerhin sind aber im untersuchten Korpus für das 19. Jh. vereinzelt Beispiele für Personenbezeichnungen in Form von Beidnennungen nachzuweisen, bspw. *Engländer und Engländerinnen*; *Leser und Leserinnen*; *Tanzkünstler und Tanzkünstlerinnen*; *Verkäufer und Verkäuferinnen* usw.: Das Zeitungs- und das Zeitschriftenkorpus enthalten 29 derartige Belege, wobei die movierte Form jeweils an zweiter Stelle steht.

Partizipkonvertat	Beleganzahl	Nomen agentis	Beleganzahl
<i>Suchende</i>	15	<i>Sucher</i>	0
<i>Schlafende</i>	7	<i>Schläfer(in)</i>	7
<i>Fahrende</i>	4	<i>Fahrer</i>	2
<i>Kämpfende</i>	4	<i>Kämpfer</i>	37
<i>Tanzende</i>	4	<i>Tän(t)zer(in)</i>	46
<i>Fragende</i>	3	<i>Frager(in)</i>	2
<i>Herrschende</i>	3	<i>Herrscher</i>	74
<i>Klagende</i>	3	<i>Kläger</i>	24
<i>Schreibende</i>	3	<i>Schreiber(in)</i>	82
<i>Mitarbeitende</i>	2	<i>Mitarbeiter</i>	29
<i>Träumende</i>	2	<i>Träumer</i>	5
<i>Lesende</i>	1	<i>Leser(in)</i>	428
<i>Mitspielende</i>	1	<i>Mitspieler</i>	2
<i>Sprechende</i>	1	<i>Sprecher(in)</i>	15
<i>Theilnehmende</i>	1	<i>Theilnehmer(in)</i>	48
<i>Zuschauende</i>	1	<i>Zuschauer(in)</i>	103

Bei elf, also mehr als der Hälfte der einschlägigen Lemmata, gibt es max. drei Instanziierungen; das Bildungsmuster scheint also nicht übermäßig produktiv, sofern Produktivität „als Häufigkeit der Regelanwendung verstanden und an der Zahl der belegten Lexeme gemessen wird“ (Demske 2000: 368). Das gilt auch, wenn man beim direkten Vergleich mit den taz-Befunden die unterschiedlichen Korpusgrößen in Rechnung stellt: Das taz-Korpus hat zwar mit seinen ca. 29,3 Mio. Wortformen den etwa siebenfachen Umfang des historischen Korpus, das nur rund 4,1 Mio. Wortformen umfasst. Doch auch bei einer Normalisierung fällt die Frequenz der relevanten Lexeme im historischen Korpus gering aus, und das quantitative Verhältnis zu den stammgleichen Nomina agentis bleibt konstant. Die Befunde deuten darauf hin, dass das Muster im historischen Korpus vor allem zu Gelegenheitsbildungen führt, aber hinsichtlich der Anwendungsdomäne nicht grundsätzlich beschränkt ist.

In einigen Fällen weicht die Semantik der Partizipialformen von der der Nomina agentis ab; so sind *Fahrende* im historischen Teilkorpus keine Personen, die ein Fahrzeug bewegen, sondern solche, die unterwegs sind. Der Stamm *Besitzende* tritt noch nicht als Kopf von Komposita auf (vgl. *Waldbesitzende*, taz 22.06.22); eine Ergänzung wird gar nicht realisiert, *Besitzende* bezeichnet also ‚Personen mit Besitz‘. Auch bei den Nomina agentis sind semantische Veränderungen zu berücksichtigen; *Mitarbeiter* wird überwiegend auf

Personen bezogen, die Beiträge zu einem Periodikum liefern; die Bedeutung ‚Beschäftigte eines Unternehmens‘ ist nur in zwei von 29 Fällen gegeben. Das Partizip *Mitarbeitende* in (70) ist direkt auf *Arbeitende* bezogen und besitzt die Bedeutung ‚Kollege‘:

- (70) Jeder Arbeitende kann übrigens in besonderen gesetzlich bestimmten Umständen, und unter Bewilligung seiner **Mitarbeitenden**, einen Urlaub erhalten. (Social-Politische Blätter, 29.08.1874)

Der Ausdruck *Studierende* ist bereits ein etabliertes Lexem (und häufiger belegt als das Äquivalent *Studenten!*); zu (*X*-)*Suchende* und (*Meist*-)*Bietende* sind, anders als nach den Wörterbucheinträgen zu erwarten, keine Nomina agentis belegt.

Zu prüfen ist nun, welcher semantische Beitrag mit denjenigen Partizipien jeweils verbunden ist, bei denen gemäß Tab. 3 durchaus ein Nomen agentis zu erwarten wäre. Die Belege (71) und (72) beziehen sich auf einen aktuellen Zustand bzw. eine zum Referenzzeitpunkt laufende Tätigkeit; es handelt sich um geradezu prototypische Beispiele für eine Gleichzeitigkeitslesart im Sinne der in Abschnitt 2.1 zitierten Grammatiken:

- (71) So beschloß er denn, eine That der Ehre und des Muthes auszuführen, und er ging hin und nahm die beiden Dolche der beiden **Schlafenden** und stach sie beide in die Brust [...]. (Morgenblatt für gebildete Leser 24.08.1856)
- (72) Dies war der Augenblick, in welchem die **Tanzenden** das Feuer zuerst bemerkten, und Aller bemächtigte sich ein panischer Schrecken [...]. Der Erzbischof, seine Dame am Arme, wollte den Uebrigen zuvorkommen und dadurch nicht nur sich und seine **Tänzerin** retten, sondern zugleich auch dem unvernünftigen Gedränge Einhalt thun. (Das Pfennig=Magazin für Belehrung und Unterhaltung 02.11.1854)

Bemerkenswert ist zudem, dass in (72) das Partizip mit einem movierten Nomen agentis alterniert, das gerade keine Gleichzeitigkeitsrelation ausdrückt, sondern eine Rolle bezeichnet; hier scheint eine klare Verteilung der Funktionen vorzuliegen.

Gleichwohl zeigen die Korpusbelege, dass bereits im 19. Jh. substantivierte Partizipien genutzt werden, die prinzipiell äquivalent mit Nomina agentis sind, auch wenn ihnen eine Gleichzeitigkeitslesart zukommt wie in (73).

- (73) 's nur die Kate, schrie da Einer oder dort, als die Tolle pfeilschnell eine Spalte niederstieg, die Wenige der **Zuschauenden** versucht haben möchten, und dann nur mit behutsamstem Fuß. (Conversations=Blatt zur Unterhaltung und Belehrung für alle Stände 31, 1837)

Diese aktuelle, aspektuell imperfektive Bedeutung wird üblicherweise durch das Nomen agentis *Zuschauer* realisiert, wie die 103 Belege zeigen; das Partizip fungiert hier offenbar als Synonym.

Gleiches gilt für *Schreibende*, *Lesende* und *Theilnehmende* (74–75); eine imperfektive Bedeutungskomponente ist irrelevant, eine Gleichzeitigkeitslesart kommt nicht in Betracht. Die Partizipien ersetzen hier die allgemein üblichen deverbale Nomina agentis.

- (74) [...] und die Ergebnisse dieser Studien mußten allmählig den Gesellschaftskörper durchdrungen haben, bevor **Schreibende** und **Lesende** da eine Fülle von Poesie entdeckten und anerkannten, wo sie eben noch die Negation aller Poesie vorausgesetzt hatten. (Morgenblatt für gebildete Leser 30.11.1856)
- (75) Die öffentliche Verhandlung des Prozesses gegen den Erzbischof von Turin, Monsignore Frasoni, hat gestern unter ungeheurem Zudrang von **Theilnehmenden** und bloßen Neugierigen stattgefunden. (Die Bayerische Presse 31.05.1850)

Festzuhalten bleibt: Substantivierte Partizipien werden bereits im 19. Jh. zur Personenreferenz benutzt. Das Muster ist insbesondere durch hochfrequente Lexeme wie *Studierende* präsent. In einigen Fällen ist eine aspektuelle Lesart im Sinne der Grammatiken anzusetzen, aber nicht alle Belege kommen für eine Gleichzeitigkeitslesart in Betracht, sie können durchaus als Varianten der Nomina agentis verstanden werden und deren Semantik aufweisen. Auch wenn angesichts der Korpusgröße eine gewisse Zurückhaltung geboten ist, legen die Befunde nahe, dass der Gebrauch substantivierter Partizipien zur Personenbezeichnung im Deutschen sprachhistorisch begründet ist und als systemadäquat gelten kann.

4. Zusammenfassung und Ausblick

Im vorliegenden Beitrag wurde der Gebrauch substantivierter Partizipien zur Personenreferenz im Vergleich mit deverbale Nomina agentis erstmals korpusbasiert untersucht. Hinsichtlich der in Abschnitt 3 skizzierten Leitfragen lässt sich festhalten:

1. In einem gegenwartssprachlichen Zeitungskorpus (taz 2000 und 2022) erweist sich dieser Wortbildungstyp als durchaus frequentes und vielfältiges Referenzmittel, wobei für die letzten 20 Jahre eine Zunahme sowohl der Lemmata als auch der Tokens zu verzeichnen ist.

2. Gleichwohl handelt es sich bei den Lemmata mit der höchsten Gebrauchsfrequenz um lexikalisierte Formen; sie nehmen im Jahrgang 2000 einen Anteil von mehr als 75 %, im Jahrgang 2022 noch mehr als 50 % der festgestellten Tokens ein.
3. Eine eingehende qualitative Analyse ausgewählter Belege hat gezeigt, dass die Partizipkonvertate im Sinne von Nomina agentis genutzt werden. Die in vielen Grammatiken und sprachkritischen Beiträgen unterstellte Gleichzeitigkeitslesart mit aspektueller Markierung liegt nur in wenigen Fällen vor. Sie ergibt sich stets aus dem Kontext und resultiert aus der jeweiligen Prädikatsklasse der verbalen Basis, nicht aus einer spezifischen Semantik des *-end*-Markers. Vor diesem Hintergrund erscheint die pauschale Kritik, die betreffenden Formen seien widersinnig oder sogar als Eingriff in die Grammatik zu werten, nicht gerechtfertigt.
4. Genutzt werden primär lexikalisierte Lemmata. Im Vergleich mit anderen Verfahren ist die Neubildung departizipialer Personenbezeichnungen von untergeordneter Bedeutung.
5. Wie der Vergleich mit Daten aus einem historischen Zeitungs- und Zeitschriftenkorpus zeigt, werden substantivierte Partizipien bereits im 19. Jahrhundert zur Personenbezeichnung genutzt und konkurrieren prinzipiell mit deverbalen Nomina agentis; es handelt sich also keineswegs um eine gegenwartssprachliche Entwicklung.

Die Befunde geben Anlass zu weiteren Fragen, denen in diesem Rahmen nicht nachgegangen werden konnte:

Unklar ist, wie das Verhältnis von substantiviertem Partizip und Nomen agentis in diachroner Perspektive zu beschreiben ist. Von einer weitergehenden Korpusanalyse zu ausgewählten Lexemen wäre Aufschluss zu gewinnen über die Entwicklung der Wortbildungsbedeutung; so stellt sich die Frage, ob vor dem 19. Jh. aspektuelle Bedeutungsmerkmale größere Relevanz besaßen und unter welchen Umständen sie ggf. abgebaut wurden.

Die Frage nach dem Prozess der Etablierung substantivierter Partizipien betrifft insbesondere auch ihre syntaktische Einbindung. Partizipien, die in erweiterten Partizipialattributen organisiert sind (*die sie Nutzenden*, vgl. 39), zeigen wegen der Beibehaltung der Argumentstellen größere Verbnahe, sind strukturell nicht mit Nomina agentis äquivalent und wurden daher in der vorliegenden Untersuchung von der Betrachtung ausgeschlossen. Der Gebrauch der Partizipien mit Possessiva oder Attributen (*ihre Nutzenden*; *Nutzende der App*) verweist wiederum auf einen hohen Nominalitätsgrad; diese Partizipien scheinen strukturell vollständig äquivalent mit deverbalen Nomina agentis. Wann und unter welchen Umständen sich diese Entwicklung vollzogen hat, ob

sie überhaupt abgeschlossen ist und inwieweit die betreffenden Wortgruppen tatsächlich akzeptiert werden, ist bislang offen.

Schließlich bleibt genauer zu untersuchen, wie sich das Verhältnis von movierten Formen, Beidnennungen, Bildungen mit Genderzeichen im Vergleich zu den Partizipien entwickelt hat und welche Tendenzen sich abzeichnen. Ob bzw. welche der genannten Formen sich in der Sprachgemeinschaft langfristig durchsetzen, bleibt ohnehin abzuwarten. Eine möglichst umfassende, linguistisch fundierte und unvoreingenommene empirische Untersuchung dieser Entwicklungen könnte weiteren Aufschluss über die Verwendung substantivierter Partizipien zur Personenreferenz bieten.

Literatur

- Adelung, Johann Christoph. 1793–1801. *Grammatisch-kritisches Wörterbuch der Hochdeutschen Mundart mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen*. 4 Bde. 2. Aufl. Leipzig: Breitkopf. ND Hildesheim & New York: Olms.
- Adler, Astrid & Albrecht Plewnia. 2019. Die Macht der großen Zahlen. Aktuelle Spracheinstellungen in Deutschland. In Ludwig Eichinger & Albrecht Plewnia (Hgg.), *Neues vom heutigen Deutsch. Empirisch – methodisch – theoretisch*, 141–162. Berlin & Boston: De Gruyter.
- Bross, Fabian. 2023. Von biertrinkenden Studierenden, schlafenden Lachenden und gendersensibler Sprache: Zur Interpretation der Gleichzeitigkeit nominalisierter Partizipien. *Sprachreport* 39(3). 40–44.
- Demske, Ulrike. 2000. Zur Geschichte der *ung*-Nominalisierung im Deutschen: Ein Wandel morphologischer Produktivität. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 122(3). 365–411.
- Diewald, Gabriele & Anja Steinhauer. 2017. *Richtig gendern. Wie Sie angemessen und verständlich schreiben*. Berlin: Dudenverlag.
- DFWB = *Deutsches Fremdwörterbuch*. Begr. v. Hans Schulz, fortgeführt von Otto Basler. Weitergeführt im Institut für deutsche Sprache. 7 Bde. Straßburg, Berlin & New York 1913–1988. 2. Aufl., völlig Neubearb. Aufl. 1995ff.
- Duden. o. J. *Duden online*. <https://www.duden.de>. (aufgerufen am 8. März 2025).
- Duden 2009. *Die Grammatik. Unentbehrlich für richtiges Deutsch*. 8., überarb. Aufl. Mannheim u. a.: Dudenverlag.
- Duden. 2016. *Die Grammatik. Unentbehrlich für richtiges Deutsch*. 9., vollst. überarb. u. erw. Aufl. Berlin: Dudenverlag.
- Duden. 2022. *Die Grammatik*. 10., völlig neu verfasste Auflage. Berlin: Dudenverlag.
- DWB = *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*. 16 Bde. in 32 Teilbänden. Leipzig 1854–1960 (ND München 1999).
- DWBN = *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*. Neubearbeitung. Hg. v. d. Berlin-Brandenburgischen Akademie der Wissenschaften und der Akademie der Wissenschaften zu Göttingen. Stuttgart & Leipzig 1965–2018.

- Eisenberg, Peter. 2007. Sprachliches Wissen im Wörterbuch der Zweifelsfälle. Über die Rekonstruktion einer Gebrauchsnorm. *Aptum. Zeitschrift für Sprachkritik und Sprachkultur* 3. 209–228.
- Eisenberg, Peter. 2020a. *Grundriss der deutschen Grammatik. Band 1: Das Wort*. Unter Mitarbeit von Nanna Fuhrhop. 5., aktualisierte und überarbeitete Auflage. Stuttgart & Weimar: Metzler.
- Eisenberg, Peter. 2020b. Die Vermeidung sprachlicher Diskriminierung im Deutschen. *Der Sprachdienst* 1–2. 15–30.
- Eisenberg, Peter. 2021. Unter dem Muff von hundert Jahren. *Frankfurter Allgemeine Zeitung* Nr. 6 (08.01.21), 12.
- Feilke, Helmuth. 2023. Gendern mit Grips statt Schreiben in Gips. Praktische Argumente für ein flexibles Gendern. *Praxis Deutsch* 297. 60–62. <https://www.friedrich-verlag.de/deutsch/sprachgebrauch/gendern-mit-grips-statt-schreiben-in-gips-sechs-thesen/> (aufgerufen am 17. Februar 2025).
- Ferstl, Evelyn C. & Damaris Nübling. 2024. Sonderzeichen als typographische Kennzeichnung geschlechtersensiblen Sprachgebrauchs. Linguistische Überlegungen und experimentelle Befunde zum Genderstern. In Sabine Krome et al. (Hrsg.), *Orthographie in Wissenschaft und Gesellschaft. Schriftsystem – Norm – Schreibgebrauch* (IDS-Jahrbuch 2023), 259–284. Berlin & Boston.
- Fleischer, Wolfgang & Irmhild Barz. 2012. *Wortbildung der deutschen Gegenwartssprache*. 4. Aufl.; völlig neu bearbeitet von Irmhild Barz unter Mitarbeit von Marianne Schröder. Berlin & New York: De Gruyter.
- Fuhrhop, Nanna & Oliver Teuber. 2000. Das Partizip 1 im Deutschen. *ZAS Papers in Linguistics* 16. 100–114.
- Glück, Helmut. 2020. Das Partizip I im Deutschen und seine Karriere als Sexusmarker. In Jessica Ammer, Gerhard Meiser & Heike Link (Hrsg.), *In vriuntschaft als es was gedâht. Freundschaftsschrift für Hans-Joachim Solms*, 381–412. Berlin: Schmidt.
- Harnisch, Rüdiger. 2016. Das generische Maskulinum schleicht zurück. Zur pragmatischen Remotivierung eines grammatischen Markers. In Andreas Bittner & Constanze Spieß (Hrsg.), *Formen und Funktionen. Morphosemantik und grammatische Konstruktion*, 159–174. Berlin & Boston: De Gruyter.
- HdR 2008 = *Handbuch der Rechtsförmlichkeit. Empfehlungen des Bundesministeriums der Justiz für die rechtsförmliche Gestaltung von Gesetzen und Rechtsverordnungen nach § 42 Absatz 4 und § 62 Absatz 2 der Gemeinsamen Geschäftsordnung der Bundesministerien*. 3., neu bearbeitete Auflage. Herausgegeben vom Bundesministerium der Justiz. Köln: Bundesanzeiger.
- Heidolph, Karl Erich, Walter Flämig & Wolfgang Motsch. 1981. *Grundzüge einer deutschen Grammatik. Von einem Autorenkollektiv*. Berlin (Ost): Akademie.
- Keller, Rudi. 2014. *Sprachwandel. Von der unsichtbaren Hand in der Sprache*. 4., unv. Aufl. Tübingen: Francke.
- Klein, Wolf Peter. 2013. Warum brauchen wir einen klaren Begriff von Standardsprachlichkeit und wie könnte er gefasst werden? In Jörg Hagemann, Wolf Peter Klein & Sven Staffeldt (Hrsg.), *Pragmatischer Standard*, 15–33. Tübingen: Stauffenburg.
- Kotthoff, Helga. 2020. Gender-Sternchen, Binnen-I oder generisches Maskulinum, ... (Akademische) Textstile der Personenreferenz als Registrierungen? *Linguistik online* 103 (03/20). 105–127.

- Kotthoff, Helga & Damaris Nübling unter Mitarbeit von Claudia Schmidt. 2024. *Genderlinguistik. Eine Einführung in Sprache, Gespräch und Geschlecht*. 2., überarbeitete u. erweiterte Aufl. Tübingen: Narr.
- Kürschner, Wilfried. 2019. Nomen Agentis vs. Partizip I. Von „Teilnehmern“ über „Teilnehmerinnen und Teilnehmer“, „TeilnehmerInnen“ zu „Teilnehmenden“. In Jürg Strässler (Hrsg.), *Sprache(n) für Europa. Mehrsprachigkeit als Chance. Auswahl an Beiträgen des 52. Linguistischen Kolloquiums in Erlangen (2017)*, 209–221. Frankfurt a.M.: Lang.
- Leibniz-Institut für Deutsche Sprache. 2024. *Deutsches Referenzkorpus / Archiv der Korpora geschriebener Gegenwartssprache 2024-I* (Release vom 13.03.2024). Mannheim: Leibniz-Institut für Deutsche Sprache. www.ids-mannheim.de/dereko (aufgerufen am 23. Juni 2024).
- Müller-Spitzer, Carolin. 2022. Gendergerechter Sprachgebrauch. Ein komplexes Feld für die Aushandlung von „richtig“ und „gut“. *Der Deutschunterricht* 74(4). 39–49.
- Müller-Spitzer, Carolin, Samira Ochs, Alexander Koplenig, Jan Oliver Rüdiger & Sascha Wolfer. 2024. Less than one percent of words would be affected by gender-inclusive language in German press texts. *Humanities and Social Sciences Communications* 11. 1343. DOI: [10.1057/s41599-024-03769-w](https://doi.org/10.1057/s41599-024-03769-w).
- Paul, Hermann. 1920. *Deutsche Grammatik*. Bd. IV. Teil IV: Syntax. (Zweite Hälfte). Halle: Niemeyer (ND Tübingen 1968).
- Payr, Fabian. *Von Menschen und Mensch*innen. 20 gute Gründe, mit dem Gendern aufzuhören*. 2. Aufl. Wiesbaden: Springer.
- Polenz, Peter v. 2008. *Deutsche Satzsemantik. Grundbegriffe des Zwischen-den-Zeilen-Lesens*. 3., unv. Aufl. Berlin & New York: De Gruyter.
- Sato, Sayaka, Ute Gabriel & Pascal M. Gygax. 2016. Altering Male-Dominant Representations: A Study on Nominalized Adjectives and Participles in First and Second Language German. *Journal of Language and Social Psychology* 35(6). 667–685.
- Scherer, Carmen. 2005. *Wortbildungswandel und Produktivität. Eine empirische Studie zur nominalen -er-Derivation im Deutschen*. Tübingen: Niemeyer.
- Schulz, Matthias. 2000. Zur Eignung von älteren Wörterbüchern und Wörterbuchvergleichen für eine Analyse von Wortschatz und dessen Wandel. *Sprachwissenschaft* 25. 63–75.
- Sökefeld, Carla. 2021. Gender(un)gerechte Personenbezeichnungen: derzeitiger Sprachgebrauch, Einflussfaktoren auf die Sprachwahl und diachrone Entwicklung. *Sprachwissenschaft* 46. 111–141.
- Stieler, Kaspar. 1691. *Der Teutschen Sprache Stammbaum und Fortwachs oder Teutscher Sprachschatz [...]*. Nürnberg: Hofmann. ND Hildesheim: Olms 1968.
- Völkening, Lena. 2022. Ist Gendern mit Glottisverschlusslaut ungrammatisch? Ein Analysevorschlag für das Suffix [ʔn] als phonologisches Wort. *Zeitschrift für Wortbildung / Journal of Word Formation* 6(1). 58–80.
- Waldendorf, Anica. 2024. Words of change: The increase of gender-inclusive language in German media. *European Sociological Review* 40. 357–374.
- Wurzel, Wolfgang Ullrich. 1988. Derivation, Flexion und Blockierung. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 41. 179–198.
- Zifonun, Gisela, Ludger Hoffmann & Bruno Strecker. 1997. *Grammatik der deutschen Sprache*. 3 Bde. Berlin & New York: de Gruyter.

Jan Seifert

Rheinische Friedrich-Wilhelms-Universität Bonn

Institut für Germanistik, Vergleichende Literatur- und Kulturwissenschaft

Rabinstr. 8

53111 Bonn

jan.seifert@uni-bonn.de



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

Melissa Mujkić

Tagungsbericht:
„Perspektiven der Movierungsforschung“
(8.–10. Oktober 2024; Universität Luxemburg)

Vom 8. bis zum 10. Oktober 2024 fand an der Universität Luxemburg am Campus Belval die Tagung „Perspektiven der Movierungsforschung“ statt, die von Caroline Döhmer (Université du Luxembourg) und Alexander Werth (Universität Passau) organisiert wurde. Sie wurde im hybriden Modus angeboten, um die Teilnahme für jene, die nicht vor Ort teilnehmen konnten, zu ermöglichen.

Im Rahmen der Tagung wurden kognitive und funktionale, diachrone, dialektale und fremdsprachige Perspektiven der Movierungsforschung präsentiert und diskutiert. Insbesondere durch die Weiterentwicklung der Korpuslinguistik und dem Zugang zu großen Sprachkorpora ist es heute möglich, diachrone und synchrone Movierungstendenzen empirisch aufzuzeigen. International aufgestellt waren nicht nur die Referent*innen, sondern auch die untersuchten Sprachen (z. B. Chinesisch, Dänisch, Deutsch, Italienisch, Luxemburgisch, Polnisch, Schwedisch), in denen moviert wird.

Den Auftakt der Movierungstagung machte **Michail K. Kotin** von der Universität Zielona Góra mit einem Vortrag zur „Movierung aus der Sicht der Opposition von System und Norm in der Sprache“. Er geht darauf ein, dass bei der Schaffung einer gendergerechten Sprache die Interaktion zwischen Sprachsystematik und Normierungsversuchen beachtet werden muss. Movierte Nomina werden als Entitäten behandelt, die danach bewertet werden, ob sie „systemkonform“ sind. Hier wird zwischen lexikalischen Paaren von diversen Stämmen (z. B. *Frau* : *Mann*) und sog. morphologisch indizierten äquipollenten Paaren (z. B. *Zauber_er* : *Zauber_in*) differenziert. In beiden

Fällen handelt es sich um reversible Oppositionen, bei denen die Zugehörigkeit der Referenten zu einem bestimmten Sexus an Suffixen erkennbar ist, die unmittelbar an den neutralen Stamm angehängt werden. Es gibt aber auch irreversible Oppositionen, bei denen der maskuline Stamm mit Nullsuffix (z. B. *Esel_* : *Esel_in*) oder vom maskulinen Stamm mit dem *er*-Suffix und seinen Erweiterungen *_ler* und *_ner* (z. B. *Lehr_er* : *Lehr_er_in*) moviert wird. Der Funktionswandel des *er*-Suffixes bei agentivischen Substantiven hat dazu geführt, dass Bedeutungen teilweise verloren gingen (Demotivierung) und eventuell neue Bedeutungen oder Funktionen angenommen wurden (Remotivierung).

Weiter ging es mit einem Vortrag von **Carolyn Müller-Spitzer** und **Samira Ochs** vom IDS Mannheim zur Frequenz movierter Personenbezeichnungen: Sie analysierten die lexikalische Varianz geschlechtsübergreifender Maskulina, Numerusverteilung, (Nicht-)Spezifität und lexikalische Verteilungen bei Movierungen. Mit dem Subkorpus DeReKo und einer Duden-Liste (10.000 Einträge) fanden sie heraus, dass Prestigebezeichnungen häufiger spezifisch maskulin und Passivbezeichnungen häufiger generisch maskulin sind. Aktiv- und Bevölkerungsbezeichnungen zeigten keine signifikanten Tendenzen. Generische Maskulina im Plural und spezifische Formen im Singular waren signifikant häufiger. Movierte Formen zeigten keine signifikanten Ausprägungen. Die häufigsten Movierungen fanden sich in der Brigitte (55,8 %), gefolgt von Zeit Wissen (18,3 %), Psychologie Heute (15,6 %) und DPA (10,3 %). 132 Movierungen (10,19 %) waren Teil einer Doppelform, der Großteil (89,81 %) stand alleine und zu 79,2 % im Singular. Müller-Spitzer und Ochs betonen die Notwendigkeit genderinklusive Sprache und empfehlen, geschlechtsübergreifende Maskulina durch genderinklusive Formen zu ersetzen. Ob die Frequenzverteilungen kognitiv im Sprachverständnis gespeichert sind, bleibt offen.

Robin Lemke, **Philipp Rauth** und **Lisa Schäfer** von der Universität des Saarlandes beschäftigten sich mit dem generischen Maskulinum, der Femininmovierung und der Anaphernresolution. Sie führten zwei Experimente durch, um herauszufinden, ob maskuline Pluralformen als sexusneutral oder sexusspezifisch interpretiert werden. In beiden Experimenten wurde ein *male bias* nachgewiesen. Verwendete Testsatzpaare enthielten allerdings Interpretationsambiguität in Bezug auf Anapher (Koreferenz vs.

Kontrast). Der *male bias* ist reduziert, wenn Ambiguität durch Prosodie (Experiment 1) oder durch einen syntaktisch kodierten Partitiv (Experiment 2) aufgelöst wird.

Dominic Schmitz (Heinrich-Heine-Universität Düsseldorf) untersuchte etablierte Formen im Kontrast zu Neografien im Bereich der Psycholinguistik. Er arbeitete mit einem Korpus (1.625.353 Sätze, 25.528.132 Token und 55.063 Types) bestehend aus Sätzen aus dem Tagesspiegel und dem Subkorpus „News“ der Leipzig Corpora Collection (Goldhahn et al. 2012). Nach der Paradigmen-Annotation hinsichtlich der Verwendung von generischen bzw. spezifischen Maskulina zeigte sich, dass generische Maskulina (11.829) am häufigsten vorkommen, gefolgt von spezifischen Maskulina (10.302), spezifischen Feminina (4.180) und Neografien (9.093). Neografien beeinflussen die Semantik etablierter Formen: Sternformen führen zu einer Rekalibrierung der Bedeutungen, wobei sie mit steigender Frequenz generischen Maskulina ähnlicher und spezifischen Maskulina unähnlicher werden. Ihre Ähnlichkeit zu spezifischen Feminina bleibt unabhängig von der Frequenz konstant. Bezüglich geschlechtsinklusive Sprache, Neografien und der Repräsentation nicht-binärer Personen bedarf es weiterer und auch innovativer Forschung, denn bislang fehlen die Ressourcen für analoge Analysen zur Repräsentation nicht-binärer Personen.

Dann präsentierten **Andreas Blombach** (FAU Erlangen-Nürnberg) und **Bettina Lindner-Bornemann** (Universität Hildesheim) die Auswirkungen und Akzeptanz gegenderter Personenbezeichnungen. In einer Fragebogenstudie erfassten sie das objektiv und subjektiv eingeschätzte Textverständnis, Formenbewertungen, Spracheinstellungen und Geschlechter- und Familienvorstellungen. Ein randomisiert zugewiesener Lokalzeitungstext mit unterschiedlichen Schreibweisen bei den Personenbezeichnungen sollte bewertet werden: generisches Maskulinum, generisches Maskulinum mit anschließend expliziter Wiederaufnahme einer Person als Vertreter:in der Gruppe, Neutralformen, Gendersternchen-Schreibung, Doppelpunkt-Schreibung. Die Ergebnisse zeigen, dass es kaum Unterschiede zwischen Schätzungen des Männer- und Frauenanteils gibt. Es gibt eine etwas geringere subjektive Verständlichkeit von Neutralformen und Sternchen. Auffälligere Unterschiede zeigen sich in den Bewertungen hinsichtlich der vorgegebenen Kategorien des Vortragenden in Bezug auf Interessantheit und Eleganz. Einen sehr deutlichen Einfluss der Einstellung zu geschlechtergerechter Sprache gibt es

bezüglich der Bewertung alternativer Formen, insbesondere zur sprachlichen Schönheit, Akzeptabilität und Regelkonformität. Zusammenhänge lassen sich zwischen Einstellung zu geschlechtergerechter Sprache und politischer Orientierung oder einzelnen Spracheinstellungen feststellen.

Lena Völkening (Carl von Ossietzky Universität Oldenburg) ging in einer synchronen konstruktionsgrammatischen Analyse auf die Funktion von Genderstern und Glottisverschlusslaut ein. Der Genderstern kann ein ikonisches Zeichen, ein graphisches Morphem oder kein Sprachzeichen sein (vgl. Zifonun 2021; Ferstl & Nübling 2024: 266–267). Das Movierungssuffix *-*in* könnte ein neues Suffix sein, für das die Sprachgemeinschaft eine neue Bedeutung aushandelt und zunehmend konventionalisiert. Dies kann konstruktionsgrammatisch erklärt werden, nämlich durch die Begegnung mit oder eigene Verwendung von vielen gegenderten Substantiven. Im Laufe der Zeit kommt es zur Generalisierung und *-*in* tritt an eine substantivische Basis. Bei dieser Modellierung ist noch die Frage offen, welche Form diese Derivationsbasis annimmt. Rauth (i. E.) zufolge wären diese Formen systemkonformer, wenn *-*in* ein „typisches“ konsonantisch anlautendes Suffix wäre, somit könnte der Stern als neues Suffix den Glottisverschlusslaut repräsentieren, indem er gegenderte Derivate von femininen Derivaten unterscheidet und damit Phonemstatus erhält. Völkening hinterfragt jedoch, ob es möglicherweise auf Dauer ein neues Suffix ist, das sich dem weiblichen Suffix annähert.

Anne Rosar von der Johannes Gutenberg-Universität Mainz stellte die Sonderstellung der Femininmovierung in Binomialen vom Typ *Bürgerinnen und Bürger* vor. In einer Korpusstudie zu den häufigsten Gender Binomialen, die alle Ausgaben von *Der Spiegel* und *Die Zeit* von 1947/1953 bis 2018 (ca. 577 Mio. Textwörter) umfasst, analysierte sie Koordinationen (*Mann und Frau*) sowie Berufs- und Funktionsbezeichnungen (*Bürger*), die Geschlecht morphologisch spezifizieren. Die dominante Abfolge wird durch eine komplexe Kombination von inner- und außersprachlichen Faktoren bestimmt, die sich gegenseitig verstärken oder abschwächen. Bei der Mehrheit der sog. Gender Binomiale steht die männliche vor der weiblichen Bezeichnung (*Mann und Frau*), also *Male First*. Dennoch gibt es eine signifikante Reduzierung des *Male First Bias* aufgrund veränderter Geschlechterrollen und Diskurse. Insofern etabliert sich in den 1980er-Jahren ein Präferenzwechsel zu *Female First Bias* bei Funktions- und Berufsbezeichnungen

(*Lehrerinnen und Lehrer*) und die Verfestigung des Musters „moviert vor unmoviert“, das vor allem sprachpolitisch induziert ist. Ob es eine Modifizierung des Sprachgebrauchsmusters durch alternierende Paarformen (*Ärztinnen und Pfleger*) geben wird, wodurch eine geschlechtsübergreifende Lesart evoziert werden soll (vgl. Ochs 2024), bleibt zum aktuellen Zeitpunkt offen.

Daniel Elmiger (Université de Genève) und **Carla Sökefeld** (Universität Hamburg) untersuchten, inwiefern eine Äquivalenz bei movierbaren Formen (M-Formen) und substantivierten Partizip-I-Formen (P-Formen) besteht. Mit Daten aus dem Frequenzwörterbuch (1 Mio. nicht-lemmatisierte Wortformen) (Quasthoff et al. 2011), Personenbezeichnungen (130 000 Wortformen) und Sprachgebrauchsdaten aus dem One Million Post Corpus (ca. 2,8 Mio Token) fanden sie Folgendes heraus: Im Frequenzwörterbuch wurden 10 212 Personenbezeichnungen mit Frequenzangabe untersucht, wovon 357 formale Entsprechungen (3,5 %) identifiziert wurden, wie *Beurteiler* und *Beurteilende*. Es zeigt sich eine Konkurrenz zwischen M-Formen und P-Formen sowie zwischen dem Gebrauch als Personenbezeichnung und anderen Verwendungsarten. In den meisten Fällen sind M-Formen frequenter als P-Formen (*Sprecher* gegenüber *(die) Sprechenden*). Wenn es um die Movierung als Alternativ- oder Konkurrenzform geht, dann zeigt sich, dass bei fast drei Viertel der maskulinen Partizipien movierte Alternativen fehlen. Wenn movierte Formen existieren, sind sie seltener als Maskulina, auch bei stereotyp weiblichen Personenbezeichnungen. M-Formen sind semantisch bereits besetzt, weisen jedoch große Bedeutungsunterschiede auf, die auf eine semantische Spezialisierung hinweisen.

Bei der vorliegenden Datengrundlage zeigt sich, dass nur bei ca. einem Drittel der P-Formen überhaupt Konkurrenz zu movierbaren Formen besteht. Allerdings bildet das Korpus nur den Zeitraum zwischen 2000 und 2015 ab, deswegen ist weitere Forschung in Bezug auf aktuelle Daten wünschenswert, um sie diachron zu vergleichen und festzustellen, welche Types bei den P-Formen dazugekommen sind.

Aline Siegenthaler der Université de Genève besprach Empfehlungen und Bewertungen sprachlicher Markierungen der Movierung in institutionellen Leitfäden für gendersensible Sprache. Ein derartiger Leitfaden soll vorsichtige bis nachdrückliche Empfehlungen zur gendersensiblen Sprache beinhalten und sprachliche Konventionen

innerhalb eines (mehr oder weniger) bestimmten Geltungsbereichs formulieren, um mehr Gendergerechtigkeit in der Sprache zu erreichen und die Leser*innen bezüglich ihres eigenen Sprachgebrauchs zu sensibilisieren (vgl. Elmiger 2021, 2022; Ivanov et al. 2018; Schlichting 1997). Neutralisierungs- und Vermeidungsstrategien werden am ehesten als unproblematisch gesehen, außer in Kombination mit binären oder nicht-binär intendierten Formen. Besonders die abgekürzten Formen (Schrägstrich, Binnen-I, Klammern) werden zusehends durch Formen mit Sonderzeichen (-x, -y, *, z. B. *dx Lesx*, *das Lesy*, *d* Les**) ersetzt. Bisher bleibt die Diskussion um die Barrierefreiheit der Formen uneindeutig, die zu unterschiedlichen Aussagen führt und nur teilweise aktuelle Studien miteinbezieht.

Adriano Murelli (Università degli Studi di Torino) diskutierte, ob und wie gendergerechte Sprache (kurz: GGS) ins DaZ-/DaF-Curriculum eingebunden werden soll. Er stellt infrage, ab welchem Sprachniveau GGS im DaF-Unterricht eine Rolle spielen sollte und ob sie Teil der elementaren Sprach- und Grammatikfertigkeiten vermittelt werden soll oder eher als Metathema dient (vgl. Hirschberg 2024: 170). Verschiedene Strategien der gendersensiblen Sprache sind möglich, doch fehlen Richtlinien für den Lernkontext, denn GGS wird in Lehrwerken weder gebraucht noch thematisiert. Aktuelle Umfragen zeigen ein Spektrum an Antworten von Ablehnung bis Akzeptanz bei Lehrenden, wodurch GGS im Lernkontext individuell angewendet wird. Aufgrund der nichtkonsequenten Verwendung von GGS in der Praxis sollen Lehrende und Lernende weiterhin entsprechend ausgebildet werden.

Sam Mersch (Zenter fir d'Lëtzebuerger Sprooch; Institut Grand-Ducal) präsentierte verschiedene Movierungsmethoden aus typologischer Sicht. Mittels des französischen Kindercomics „Les Pyjamasques“ von Romuald Racioppo zeigte er, wie der Name der Protagonistin *Owlette* in verschiedenen Sprachen übersetzt wird, um verschiedene Movierungsstrategien aufzuzeigen: Morphologische Markierung (*owlette* von ‚owl‘ im Englischen), lexikalische Markierung (*maō tóu yīng nǚ* von ‚Eule + Frau‘ im Chinesischen), keine Markierung (*Bagoly* ‚Eule‘ im Ungarischen) oder Entlehnung (*Аулет* [aulet] im Bulgarischen). Weiterhin beschrieb Mersch die verschiedenen Markierungsmöglichkeiten im Spanischen, Koreanischen, Französischen, Chinesischen und Japanischen, um die Unterschiede in den jeweiligen Sprachen anhand von Beispielen aufzuzeigen.

Jakob Böhm (Universität Leipzig) beschäftigte sich in einer explorativen Korpusstudie mit dem Konzept der Movierung im Althochdeutschen. Böhms Hypothesen zielen darauf ab, dass die Genusdifferenzierung der primäre Prozess der Movierung ist. Erst sekundär entwickelt sich die Derivation als hauptsächlichlicher alleiniger Movierungsprozess im Deutschen. Durch den morphologischen Wandel von der Genusdifferenzierung zur Derivation im Althochdeutschen erfährt die Movierung eine funktionale Einengung von der Sexusspezifizierung hin zur Weiblichkeitsspezififizierung. Das althochdeutsche *-ârin* schlägt sich in der Überlieferung hauptsächlich erst in der Periode des Mittelhochdeutschen nieder. Das Suffix *-âra* wird im Ahd. von *-ârin* verdrängt (z. B. *klag-âra* > *klagâr-in* ‚Klägerin‘). Diese Entwicklung der Movierung ist eingebettet in eine Transformation des Nominalwortbildungssystems von einem Stammbasissystem (Genusdifferenzierung) zu einem Wortbasissystem (Derivation).

Claudia Wich-Reif (Rheinische Friedrich-Wilhelms-Universität Bonn) demonstrierte, wie Movierungen allgemein und im Herkunfts-Duden aufgenommen werden, die bisher eher unsystematisch erfasst wurden. Heute gibt es das Bestreben, die entsprechenden Formen möglichst umfassend zu dokumentieren. Inhaltlich-formal gibt es bei der Aufnahme insofern Unterschiede, als es vor allem für weibliche Personen- und Berufsbezeichnungen vollständige Artikel oder bloß Verweisartikel gibt. Movierungen im Herkunfts-Duden werden im Vergleich zur Vorgängerauflage systematischer aufgenommen, und zwar sowohl in quantitativer Hinsicht wie auch in Bezug auf die zeitlichen Dimensionen.

Trotz des begrenzten Umfangs des Herkunfts-Duden als gedrucktes Wörterbuch spiegeln die Movierungen, insbesondere mit der *in*-Endung, die außersprachlichen Entwicklungen gut wider. Belege aus Texten des 18. und 19. Jahrhunderts dokumentieren die Notwendigkeit, weibliche Personen sprachlich sichtbar zu machen. Auch **Eva Bütthe-Scheider** (Sächsische Akademie der Wissenschaften zu Leipzig) äußerte sich zur (Neu-)Einführung von Movierungen und äquivalenten Kompositionsbildungen in Enzyklopädien (sog. „braune Meyer“ 1936–1942) mit Fokus auf der Zeit des Nationalsozialismus. Von Auflage zu Auflage wird eine Veränderung bezüglich der Anzahl und der Art der Lemmaansätze festgestellt (z. B. von *Frauenrechtler* (6. Aufl. 1904) über das Verweislemma *Frauenrechtlerin* (7. Aufl. 1926/1933) zum Vollartikel

Frauenrechtlerin (8. Aufl. 1938)). Movierungen oder äquivalente Komposita werden in den Enzyklopädien konsequent mittels *-in* ausgedrückt (außer *Hebamme*). Mit der zahlreichen Verwendung von *-in* geht jedoch kein Fortschritt in der Emanzipation einher.

Hanna Augustin (Universität Passau) untersuchte die onymische Movierung im Mittelbairischen zu ihrer phonologischen und pragmatischen Lizenzierung in vier Regionen (Seibersdorf, Herrenau, Kirchenegg und Seeg). Sie wollte herausfinden, welchen phonologischen Beschränkungen die onymische Movierung unterliegt und wann sie verwendet wird. Die pragmatischen Funktionen der onymischen Movierung standen ebenfalls im Fokus: Art der Bezugnahme (adressierend, referierend), Markierung von Nähe oder Distanz, Auf- oder Abwertung. Abgefragt wurden 144 Familiennamen nach ihrer Movierbarkeit im Bairischen und Schwäbischen. In Fotogesprächen sollte eine möglichst natürliche Sprache aufgenommen werden. In Einzelinterviews wurde eine metasprachliche Einschätzung eingeholt, indem Fragen zu typischen Verwendungssituationen, Wirkung und Funktionen der Movierung gestellt wurden. Es wurde auch erfragt, für welche Personen die Movierung verwendet wird und mit welchen Gesprächspartner*innen sie angewendet wird. Die onymische Movierung ist nie obligatorisch, dennoch gibt es spezielle Situationen, in denen eher moviert wird. Unterschiede gibt es zwischen dem Bairischen (eher gute Bekannte) und dem Schwäbischen (auch Frauen in der Distanz). Die onymische Movierung ist ein nächstes sprachliches Phänomen, das v. a. von älteren ortsgebundenen Personen für die Referenzierung auf ältere ortsgebundene Frauen verwendet wird.

Alexander Werth der Universität Passau befasste sich mit Movierungsfunktionen durch Komposition in allen belletristischen Texten des Deutschen Textarchiv im DWDS von 1465 bis 1987. Um Ähnlichkeiten aufzuzeigen, erfolgte die Suche nach *-tochter*, *-frau* und *-in* mit entsprechenden Schreibvarianten im Singular (vgl. Stephan 2009: 359). Vorherhige Studien (s. Haß-Zumkehr 2003; Eisermann 2003; Stephan 2009) haben bereits gezeigt, dass *-frau* seltener vorkommt als *-in*. Ob *-frau* eine funktionale oder matrimoniales Lesart hat, ist historisch stark abhängig vom Status des Erstglieds. Eine matrimoniales Lesart wird bei belebten Erstgliedern evoziert, eine funktionale hingegen bei unbelebten. Die Befunde zeigen, dass *-frau* bis ins 20. Jh. nur schwach produktiv ist, was mit dem weiten Funktionsspektrum von *-in* und mit dem generellen Abbau der matrimonialen

Lesart im Zusammenhang steht. Zudem ist „bei den Komposita mit dem Zweitglied *-frau* [...] die matrimonielle Bedeutung stärker als bei den Komposita mit dem Zweitglied *-weib*“ (Stephan 2009: 363).

Caroline Döhmer und **Alistair Plum** (Université du Luxembourg) fragten sich, wie Frauen in Online-Kommentaren repräsentiert sind. Im Luxemburgischen wird zwischen 11 Movierungstypen unterschieden: *-in*, *-esch*, *deadj.*, *-fra*, *-e*, *-rice*, *-ière*, *-euse*, *-esse*, *epizön* und weitere. Dabei untersuchen sie die häufig auftretenden Formen und welche Funktion Referenzen auf Frauen haben. Das Korpus, mit dem sie arbeiteten, umfasst ca. 1 Mio. Online-Kommentare (50 Mio. Tokens) zu Nachrichtenartikeln auf RTL.lu von 2009 bis 2023. Feminine Movierungsmuster kommen häufig bei Frauen als Führungspersonen oder Amtsinhaberinnen (z. B. *Ministesch* ‚Ministerin‘), Stellvertreterinnen eines Berufs (z. B. *Infirmière* ‚Krankenpflegerin‘) und als Teil einer Beidnennung (z. B. *Bierger a Biergerinnen* oder *Biergerinnen a Bierger* ‚Bürger und Bürgerinnen‘ oder ‚Bürgerinnen und Bürger‘) vor. Die *in*-Movierungen sind dominant und weibliche Personenbezeichnungen zeigen sich vor allem bei „systemrelevanten Berufen“ bzw. Berufen, die in typischer Frauenhand sind.

Damaris Nübling (Johannes Gutenberg-Universität Mainz) untersuchte, wo die schwedische Movierung steht und wie es um ihre (Un-)Produktivität bestellt ist. Im Schwedischen gibt es zwei Movierungssuffixe: *-inna* und *-(er)ska*. Das Suffix *-inna* wird eher für zufällige, temporäre Nebenrollen (*beundrarinna* ‚Bewunderin‘) verwendet und *-(er)ska* eher für einfachere Berufe (*städerska* ‚Reinigungsfrau‘) und Langzeittätigkeiten (vgl. Wessén 1992). In früheren matrimoniellen Verwendungen (*generalska*) und heutigen Neubildungen (*bloggerska* ‚Bloggerin‘) sind diese Movierungssuffixe präsent. Eine semantische Beschränkung gibt es bei nicht-movierbaren statushohen Berufen (*läkare* ‚Arzt‘). Teilweise gibt es auch Pejorierungen (*lärarinna* ‚ältliche Grundschullehrerin‘). Heute werden Movierungen von der Sprachbehörde, die seit den 1960er-Jahren von Movierungen abrät, als unproduktiv, stigmatisiert, degradierend und sexistisch erklärt (vgl. Språkriktighetsboken 2005: 81). Trotz alledem gibt es weiterhin echte, nicht pejorative Neubildungen (Typ *prisutdelerska*) bzw. produktive Weiterverwendung movierter Bausteine (Typ *rock-*, *wagnersångerska*). Das Suffix *-(er)ska* ist produktiver als *-inna*.

Mirjam Schmuck (Universität Kopenhagen) stellte die Femininmovierung im Dänischen vor, die mit den Suffixen *-inde* und *-ske* gebildet wird. Das Suffix *-ske* ist älter (*syerske* ‚Näherin‘) und wird auch matrimoniell bei Familiennamen (*Møllerske*) verwendet. Ansonsten wird es meist durch das prestigehaltigere *-inde* (*sangerske* > *sangerinde* ‚Sängerin‘) ersetzt. Das Suffix *-inde* hat weniger formale Beschränkungen als *-ske*, v. a. bei Amts- und Standesbezeichnungen sowie bei Berufen der Mittel- und Oberschicht. Teils wird es matrimoniell (*professorinde* ‚Professorin‘) und teils funktional (*forstanderinde* ‚Vorsteherin‘) verwendet. Ende des 19. bzw. Anfang des 20. Jh. wird *-ske* durch *-inde* ersetzt (*lærerske* > *lærerinde* ‚Lehrerin‘). Seit dem Rückgang movierter Formen ab Mitte des 20. Jahrhunderts gelten beide Suffixe heute als unproduktiv.

Schmuck hat eine diachrone Untersuchung mit Wörterbuchdaten (3,4 Mrd. Tokens) durchgeführt, um höhere Amts- und Standesbezeichnungen, traditionelle Frauenberufe, Künstlerinnen, stereotyp-weibliche Rollen und andere (z. B. spontane Rollen) zu untersuchen und kommt zum Ergebnis, dass im Dänischen nur noch Movierungsreste vorhanden sind. Die Markierung der Weiblichkeit (Genuskongruenz) zur Markierung weiblicher Stereotypen (*lærarinde* ‚weiblicher Lehrer‘ meint hiermit einen minderwertigen Lehrer bzw. Hobby-Lehrer) führt letztlich zur Vermeidung und zum Abbau.

Hanka Błaszowska (Adam-Mickiewicz-Universität Poznań) widmete ihren Vortrag der Femininmovierung im Polnischen. Weniger frequente Suffixe werden für neue Verwendungen aktiviert: *redaktora* ‚Redakteurin‘, *pankówna* ‚Punkerin‘, *jesieniara* ‚Herbstliebhaberin‘. Die Produktivität zeigt sich am Movierungssuffix *-ka*, das auch bei fremdsprachigen Wortbildungsbasen eingesetzt wird (*fajterka* ‚Fighterin‘), am Suffix *-ini/-yni* (*sekretarzyni* ‚Sekräterin‘) und an der paradigmatischen Ableitung auf *-a* (*dziekana* ‚Dekanin‘). Die Zahl der neuen Derivate wächst, parallel dazu werden auch traditionelle Derivate wiederbelebt, die eine stilistisch-semantiche Umwertung erfahren (*gościni* ‚Gästin‘). Konsequenzen für den allgemeinen Sprachgebrauch gibt es insofern, dass movierte Feminina weit verbreitet und neue Movierungstendenzen ersichtlich sind, wie etwa in der Politik, in den Medien und in Wörterbüchern.

Valentina Crestani (Università degli Studi di Milano) zog die deutsche ‚leichte Sprache‘ und italienische ‚lingua facile‘ heran, um sie mittels zwei Korpora zu vergleichen. Sie zeigte, dass das Deutsche bezüglich Normiertheit vorangeschrittener ist als das

Italienische, das sich noch im Anfangsstadium befindet. Unterschiede gibt es auch bezüglich des Aufgreifens der gendersensiblen Sprache: Im Italienischen wird sie nicht explizit thematisiert, wohingegen deutsche Regelwerke die Movierung miteinbeziehen. Im Vergleichbaren Korpus spielt die Movierung eine untergeordnete Rolle, denn Feminina mit weiblicher Referenz und Doppelnennungen weisen niedrige Prozentsätze auf. Im Parallelkorpus zeigt sich, dass die Verwendung einer geschlechtersensiblen Sprache in deutschen Leichte-Sprache-Texten einen Einfluss auf die Wiedergabe von Texten im Italienischen hat. Für fast alle untersuchten Kategorien gilt, dass der höchste Prozentsatz der Formen im Italienischen der jeweiligen Kategorie im Deutschen entspricht. Die einzige Ausnahme stellen die Substantive mit dem Gendersternchen und mit dem Gendergap dar. Feminina, Doppelnennungen und Gendersternchen-Formen sind Kategorien mit niedrigen Prozentsätzen.

Zum Schluss der Movierungstagung fand eine abschließende, intensive Diskussion über die behandelten Themen, erzielten Ergebnisse und potenziellen zukünftigen Forschungsrichtungen statt. Aus den Vorträgen ergab sich ein bereichernder Austausch zu verschiedenen Wortbildungsmustern (z. B. Wandel, Restriktion, Kreativität, Präferenz) und diatopischer Variation sowie zu unterschiedlichen Funktionsbereichen und -unterschieden. Insgesamt behält die Movierung ihre Relevanz in der (korpusbasierten) Forschung sowohl aus linguistischer als auch aus gesellschaftlicher Perspektive und soll weiterhin Gegenstand zukünftiger Forschung sein.

Literatur

- Eisermann, Sonja Iris. 2003. Berufsbezeichnungen für Frauen vom 16.-19. Jh. Diss. Oldenburg.
- Elmiger, Daniel. 2021. Les guides de langue non sexiste / inclusive dans les langues romanes. Un genre textuel évolutif. In Benjamin Fagard & Gabrielle Le Tallec (Hrsgg.), *Entre masculin et féminin. français et langues romanes*, 119–136. Paris: Presses Sorbonne Nouvelle.
- Elmiger, Daniel. 2022. Leitfäden für geschlechtergerechte Sprache im Verlauf der Zeit. Tendenzen in den romanischen Sprachen. In Lydia Becker, Julia Kuhn, Christina Ossenkop, Claudia Polzin-Haumann & Elton Prifti (Hrsgg.), *Romanistisches Kolloquium XXXV. Geschlecht und Sprache in der Romania. Stand und Perspektiven*, 67–106. Tübingen: Narr Francke Attempto.
- Ferstl, Evelyn & Damaris Nübling. 2024. Sonderzeichen als typographische Kennzeichnung geschlechtersensiblen Sprachgebrauchs. Linguistische Überlegungen und experimentelle Befunde zum Genderstern. In Sabine Krome, Mechthild Habermann, Henning Lobin &

- Angelka Wöllstein (Hrsgg.), *Orthographie in Wissenschaft und Gesellschaft. Schriftsystem – Norm – Schreibgebrauch*, 259–284. Berlin & Boston: De Gruyter.
- Haß-Zumkehr, Ulrike. 2003. Hat die Frauenbewegung Wortschatzgeschichte geschrieben? In Martin Wengeler (Hrsg.), *Deutsche Sprachgeschichte nach 1945. Diskurs- und kulturgeschichtliche Perspektiven. Beiträge zu einer Tagung anlässlich der Emeritierung Georg Stötzels* (Germanistische Linguistik 169–170), 161–179. Hildesheim, Zürich & New York: Olms.
- Hirschberg, Tim. 2024. Geschlechtergerechte Sprache und Deutsch als Fremdsprache (DaF). In Ewa Trutkowski & André Meinunger (Hrsgg.), *Gendern – auf Teufel*in kommt raus?*, 168–177. Berlin: Kulturverlag Kadmos.
- Ivanov, Christine, Maria Lange & Tabea Tiemeyer. 2018. Geschlechtergerechte Personenbezeichnungen in deutscher Wissenschaftssprache. Von frühen feministischen Vorschlägen für geschlechtergerechte Sprache zu deren Umsetzung in wissenschaftlichen Abstracts. *Suvremena lingvistika* 44(86), 261–90.
- Quasthoff, Uwe, Sabine Fiedler & Erla Hallsteinsdóttir. 2011. *Frequency Dictionary German/Häufigkeitswörterbuch Deutsch*. Leipzig: Leipziger Universitätsverlag.
- Rauth, Philipp (i. E.). Freund:innenliche Chef:innen. Eine Korpusstudie zu Frequenz und Paradigmatizität von Gendermovierungen. In Alexander Werth (Hrsg.), *Die Movierung. Formen, Funktionen, Bewertungen*, Berlin & Boston: de Gruyter Brill (Empirische Linguistik).
- Schlichting, Dieter. 1997. Nicht-sexistischer Sprachgebrauch. Über Sprachratgeber für kommunikative Zwickmühlen. *Sprachreport – Informationen und Meinungen zur deutschen Sprache* 2, 6–11.
- Språkriktighetsboken. 2005. *En handbok i svenska språket*. Stockholm: Norstedts.
- Stephan, Julia. 2009. Wortbildungsmodelle für Frauenbezeichnungen. Hamburg.
- Ochs, Samira. 2024. Von Ärztinnen und Pflegern. Eine korpusbasierte Untersuchung zu neuen Arten der Beidnennung in Die Zeit. Vortrag auf der GAL-Jahrestagung, Sektion Medienkommunikation. 12.9.2024, Dresden.
- Wessén, Elias. 1992. Svensk språkhistoria: Ordbildningslära. Akademitryck.
- Zifonun, Gisela. 2021. Eine Linguistin denkt nach über den Genderstern. *Sprachreport* 37 (2), 46–51.

Melissa Mujkić

Universität Luxemburg

Institut für luxemburgische Sprach- und Literaturwissenschaft

11, Porte des Sciences

4366 Esch-sur-Alzette, Luxemburg

melissa.mujkic@uni.lu



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