

Borrowed Approximative Morphemes in Finnish Collocations

Daniel Ebner, Humboldt-Universität zu Berlin, Germany

Abstract

This study presents a corpus-based account of the usage of three borrowed morphemes in Finnish collocations: feikki, kvasi- and pseudo-. All three can be identified as loans, based on their phonetic and phonotactic structure. Their semantic profiles overlap according to dictionary definitions, and when combining with independent lexical items they may function as categorizing and evaluating elements of the base they combine with. The results of this study, however, show that the overlap among these three morphemes is rather restricted: the morphemes only share a small number of collocate types in the analysed samples and most of the attracted collexemes identified using a Multiple Distinctive Collexeme analysis correspond to frequent collocate types of each morpheme. In the analysis, pseudo- often occupies an intermediate position between feikki and kvasi-, e.g., with regard to the part of speech of the collocate or productivity scores. Still, the overlap between kvasi- and pseudo- seems to be larger than between feikki and either of the other two morphemes.

Keywords: approximation, distinctive collexeme analysis, evaluative morphology, Finnish, productivity.

1 Introduction

This study examines the use of three borrowed morphemes (*feikki*, *kvasi-* and *pseudo-*) in Finnish. When combined with other elements, cf. examples (1)–(3), the resulting collocations differ in meaning from the simple base in (often intersecting) ways relating to broader concepts such as similarity, (non-)prototypicality and evaluation.

- (1) *ja nyt “feikki-crocsit” ja lenkkarit ovat [...] järjestyksessä*
 and now fake-croc:NOM;PL and sneaker:NOM;PL be.3PL order:INESS
 ‘and now the **“fake Crocs”** and sneakers are [...] well arranged’
- (2) *Katsaukseen hyväksyttiin satunnaistetut ja*
 review:ILL accept:PST;PASS randomize:PTCP.NOM;PL and
kvasisatunnaistetut tutkimukset
 kvasi.randomize:PTCP.NOM;PL study:NOM;PL
 ‘Randomized and **quasi-randomized** studies were accepted for the review’
- (3) *Pseudotiede taas on yleisesti aina epätiedettä*
 pseudo.science[NOM] again be.3SG general:ADV always NEG:science:PART
 ‘**Pseudoscience**, on the other hand, is generally always non-science’

Though the ‘fake Crocs’ of example (1) refer to certain properties of the shoe brand Crocs (similarity), they do not belong (at least: not undisputedly) to the category of Crocs shoes (and are thus non-prototypical examples of Crocs shoes). In addition, the collocation can be thought of as expressing an (individual) assessment of either coming close to a prototype, in this case the ‘authentic’ Crocs, or falling short of it (evaluation). These semantic modifications may be subsumed under the term APPROXIMATION, a term that covers a variety of values, among them

privativity (or category exclusion), subjective depreciation, attenuation, fakeness and similarity. All these values relate to the expression of a deviation from a standard/norm/prototype – i.e., in the examples above, a deviation from ‘standard/real Crocs’ in (1), ‘standard/normal randomization’ in (2) and ‘real/prototypical science’ in (3). The degree of deviation ranges from category exclusion to peripheral inclusion in the standard/norm/prototype (for an exhaustive list of values and the relation of approximation to categorization, cf. Masini et al. 2023: 10–16). Still, according to Schlücker (2026: 885), the values just mentioned “have in common [...] that they express a certain degree of resemblance or proximity between the base unit X of the approximative expression and the concept Y which is denoted by the approximative expression”.

Approximation and evaluation are closely related. Schlücker (2026: 884) refers to approximation and evaluation as “two distinct but closely intertwined concepts” and points out that the hierarchy between the two is not always clear, as evaluative morphology is treated in the literature either as a subclass or superclass of approximation. In their model of evaluative morphology, Grandi & Körtvélyessy (2015: 11–12) opt for the latter, placing the notion of approximation/reduction/attenuation at the intersection of “qualitative perspective” (indicating interpretation rather than description) and “shift towards the negative end” (indicating negative feelings of the speaker). In a more abstract and general version of their model, this very intersection is termed “prototypical pejoratives”, thus linking approximation/reduction/attenuation to pejoration (Grandi & Körtvélyessy 2015: 12; on pejoration in general, see Finkbeiner et al. 2016). This link is also reflected by the inclusion of the value ‘subjective depreciation’ within the cover term ‘approximation’ in Masini et al. (2023: 11).

The three morphemes considered in this study, *feikki*, *kvasi*- and *pseudo*-, have been discussed in the context of approximation and/or evaluative morphology for a number of languages, with some studies also considering other “fake” or approximative morphemes alongside (cf. Van Goethem & Norde 2020 for *fake* in Dutch; Cappelle et al. 2023 for *pseudo*- and *quasi*- in English; Vassiliadou et al. 2023 for *pseudo*(-) in Greek and French; Van Goethem et al. 2025 for *pseudo*- in Germanic and Romance languages).¹

This study aims to contribute to the discussion of the morphemes from the perspective of a typologically different language. While Finnish dictionary definitions suggest semantic overlap for the three morphemes, it is not clear to what extent this overlap is reflected in actual language use. The present study therefore adopts a corpus-based approach to investigate whether this semantic overlap is reflected in actual usage, focusing on segmentable collocational constructions in Finnish.

The outline of the paper is as follows: After a description of the codification of the morphemes in dictionaries and an overview on their morphological properties in section 2, section 3 provides the research questions for the present study, section 4 is concerned with data and methods and section 5 presents the results to be discussed in section 6. Section 7 contains the conclusions as well as suggestions for possible future research.

¹ Results of these studies will be discussed in more detail in §6.3.

2 Description of the morphemes

2.1 Codification in dictionaries, etymology and shared definitions

The Dictionary of Contemporary Finnish (*Kielitoimiston sanakirja*) and the Finnish Etymological Dictionary (*Suomen etymologinen sanakirja*)² list separate entries for *feikki*, *kvasi*-, and *pseudo*-.

*Feikki*³ is categorised as both an adjective and a noun. The definitions provided are *väärennetty* ‘forged’ and *epäaito* ‘ungenuine’ for adjectival use and *väärennös* ‘forgery’ for nominal use. It is classified as *arkikieli* ‘informal language’. Related lexical Finnish items listed in the dictionary are the verb *feikata* ‘to fake’ and its nominal derivations *feikkari* ‘faker’ and *feikkaus* ‘faking’.

Kvasi–⁴ and *pseudo*–⁵ are both described as “yhdyssanojen alkuosa[...]”, i.e., ‘first part of a compound’, a claim strengthened by the inclusion of a hyphen at the end of the respective lemma. The definitions provided for *kvasi*- and *pseudo*- include both (free) adjectives and bound forms, the latter indicated by a hyphen (such as *näennäinen*, *näennäis*- ‘ostensible’). The dictionary also lists separate lemma entries for words whose initial element is *kvasi*- or *pseudo*- (*kvasifilosofinen* ‘quasi-philosophical’, *kvasitieteellinen* ‘quasi-scientific’, *pseudoartroosi* ‘pseudoarthrosis’, *pseudokruppi* ‘pseudocroup’ and *pseudonyymi* ‘pseudonym’).

Etymologically, *feikki* can be traced to English *fake*,⁶ which is listed in the Oxford English Dictionary both as a noun and an adjective,⁷ while *kvasi*- and *pseudo*- can be traced back to respectively Latin *quasi* ‘as if’⁸ and Greek *ψευδο*- (which ultimately derives from the Verb *ψεύδω* ‘to lie’).⁹ All three morphemes are identifiable as loans in Finnish due to their phonetic and phonotactic structure, as the phoneme /f/ (found in *feikki*) “only occurs in relatively recent loanwords” (Suomi et al. 2008: 35) and the initial consonant clusters found in *kvasi*- and *pseudo*- are indicative of loans as well (cf., e.g., Karlsson 2017: 22). Concerning orthographic nativization, *pseudo*- only shows – not surprisingly – the use of the Latin rather than the Greek alphabet, while *feikki* and *kvasi*- have been adapted more visibly to fit the Standard Finnish orthographic system. In the case of *feikki*, the final vowel *-i* integrates the morpheme into the Finnish inflectional system (cf., e.g., Battarbee 2004: 270).

The definitions provided in the Dictionary of Contemporary Finnish show substantial overlap between the three morphemes, suggesting a similar semantic profile (see Table 1): While *väärä* ‘false’ is used to describe *pseudo*-, the etymologically related terms *väärennetty* ‘forged’ and *väärennös* ‘forgery’ (both derived from the verb *väärentää* ‘to counterfeit’, lit. ‘to make false’) are used to define *feikki*; *epäaito* ‘ungenuine’ is used as a definition for both

² Accessible, respectively, online at <https://www.kielitoimistonsanakirja.fi/> and <https://kaino.kotus.fi/ses/>.

³ Kielitoimiston sanakirja, s.v. “feikki”, October 2025 (<https://www.kielitoimistonsanakirja.fi/feikki>).

⁴ Kielitoimiston sanakirja, s.v. “kvasi-”, October 2025 (<https://www.kielitoimistonsanakirja.fi/kvasi->).

⁵ Kielitoimiston sanakirja, s.v. “pseudo-”, October 2025 (<https://www.kielitoimistonsanakirja.fi/pseudo->).

⁶ Suomen etymologinen sanakirja, s.v. “feikki”, October 2025

(https://kaino.kotus.fi/ses/?p=article&etym_id=ETYM_0b420c67a81dd82fa43b69900f518580&word=feikki).

⁷ Oxford English Dictionary, s.v. “fake (n² & adj.)”, October 2025 (<https://doi.org/10.1093/OED/9975759435>).

⁸ Suomen etymologinen sanakirja, s.v. “kvasi-”, October 2025

(https://kaino.kotus.fi/ses/?p=article&etym_id=ETYM_ecebc288f6cb327b7b9bfc8cec79937f&word=kvasi-).

⁹ Suomen etymologinen sanakirja, s.v. “pseudo-”, October 2025

(https://kaino.kotus.fi/ses/?p=article&etym_id=ETYM_210fe16f1e71113a57bd84169fa64c94&word=pseudo-).

(adjectival) *feikki* and *pseudo-*; *näennäinen* ‘ostensible’ is used to define both *kvasi-* and *pseudo-*; and, finally, one of the definitions given for *pseudo-* is *kvasi-*.

Table 1: Overlap in dictionary definitions; x indicates usage in the definition of the lemma

	<i>väärä</i> ‘false’	<i>epäaito</i> ‘ungenuine’	<i>näennäinen</i> ‘ostensible’	<i>kvasi-</i> ‘quasi’
<i>feikki</i> (N)	<i>väärennös</i> ‘forgery’			
<i>feikki</i> (A)	<i>väärennetty</i> ‘forged’	x		
<i>kvasi-</i>			x	
<i>pseudo-</i>	x	x	x	x

The overlap in meaning may also account for the juxtaposition of these morphemes (as in *kvasi/pseudouskonnollisessa proosarunoelmassaan* ‘in his *kvasi/pseudoreligious* prose poem’ [fiTenTen24]), where the slash could be interpreted as an indication that *kvasiuskonnollinen* ‘quasi-religious’ and *pseudouskonnollinen* ‘pseudo-religious’ may be treated as synonymous.

The data used for this study contain bases that combine with more than one of the three morphemes *feikki*, *kvasi-* and *pseudo-*. This could be taken to indicate that the three morphemes are in competition, either between individual words or collocations that contain them or, on a more abstract level, between word-formation patterns (cf. Bauer 2006: 181). I return to the issue of shared bases in §5.4.

2.2 Morphological properties

Although there are currently no studies on the diachronic development of these three morphemes, the dictionary classifications of *feikki* as both an adjective and a noun, and of *kvasi-* and *pseudo-* as the first elements of compounds, provide indirect evidence of their predominant usage in contemporary Finnish. These definitions may also reflect how they entered the Finnish lexicon: i.e., *feikki* as an independent lexical item and *kvasi-* and *pseudo-* as bound items. Yet corpus data show that the three items are used in other ways as well.

All three morphemes can act as the base of derivational formations, as in the noun *feikkiys* ‘fakeness’ (*feikki-ys* fake-DER, cf. (4)), the verb *pseudoilla* ‘to [draw in a] pseudo [style]’ (*pseudo-il-la* pseudo-DER-INF, cf. (5)) or the adjective *kvasimainen* ‘quasi-like’ (*kvasi-mainen* quasi-DER, cf. (6)):

- (4) *ja kyllä sen feikkiydenkin tunnistaa*
 and yes it:GEN fake:DER:GEN:CLT recognise:3SG
 ‘and yes one can recognise its **fakeness**’ (fiTenTen24)
- (5) *pitää osata muutakin kuin pseudoilla.*
 must:3SG be.able:INF other:PART:CLT than pseudo:DER:INF
 ‘you have to know more than **to [draw in a] pseudo [style]**.’ (fiTenTen24)
- (6) *ne vaikuttavat kaikkein kotikutoisimmilta ja kvasimaisilta.*
 3PL seem:3PL all:SUP homemade:SUP:PL:ABL and kvasi:DER:PL:ABL
 ‘they seem the most homespun and **quasi-like**.’ (fiTenTen24)

Forms corresponding formally to all three morphemes may also occur independently, either as nouns (7), (inflected) adjectives (8), underspecified elements¹⁰ (9) or sentence adverbials (10).

- (7) *Annetaanpa pseudolle mahdollisuus*
 give:PASS:CLT pseudo:ALL chance[NOM]
 ‘Let’s give **the pseudo** a chance’ (fiTenTen24)
- (8) *Tai onkohan hampaat ylipäättään aidot vai feikit?*
 or be.3SG:CLT tooth:NOM;PL generally real:NOM;PL or fake:NOM;PL
 ‘Or are the teeth in general real or **fake**?’ (fiTenTen24)
- (9) *Näyttää feikiltä tossa kuvassa*
 appear:3SG fake:ABL this:INESS picture:INESS
 ‘It looks [**like a**] **fake** in this picture’ (fiTenTen24)
- (10) *kuten sinä teet kvasi*
 like 2SG do:2SG kvasi
 ‘like you do **quasi**’ (fiTenTen24)

Finally, all three morphemes are found as initial elements in collocational constructions, where they are part of a collocation that may or may not be segmentable into independently occurring lexical items synchronically.¹¹ This specific use and their collocational behaviour will be the focus of sections 5 and 6. Non-segmentable collocations are primarily neoclassical formations as in (11); in the case of segmentable collocations, *feikki*, *kvasi*- and *pseudo*- can combine with elements pertaining to different parts of speech (12)–(14).

- (11) [non-segmentable]
Barbara Mertz kirjoitti romaaninsa pseudonyymeilla
 Barbara Mertz write:PST[3SG] novel:POSS3 pseudonym:PL:ADESS
 ‘Barbara Mertz wrote her novels under **pseudonyms**’ (fiTenTen24)
- (12) [segmentable; base: proper noun]
joo multakin löytyy parit tollaset feikkiuggit
 yeah 1SG:ABL:CLT be.found:3SG pair:NOM;PL such:NOM;PL fake.ugg:NOM;PL
 ‘yeah, I too have a pair of these **fake Uggs**’ (fiTenTen24)
- (13) [segmentable; base: noun]
Liikutaan jo tieteen ja pseudotieteen rajoilla...
 move:PASS alreadyscience:GEN and pseudo.science:GEN border:PL:ADESS
 ‘We are already moving on the borderline between science and **pseudoscience**...’
 (fiTenTen24)

¹⁰ ‘Underspecified’ means that the element can be interpreted either as a noun or as an adjective. Certain slots in Finnish constructions can be filled by both nouns and adjectives, which might make it difficult to distinguish between the two parts of speech. Examples are predicative use and certain constructions such as *näyttää X-ltä* ‘looks (like) X’ (where the X slot can be filled with a typical noun, such as *isä* ‘father’ in *näyttää isältä* ‘looks like father’, or a typical adjective, such as *hyvä* ‘good’ in *näyttää hyvältä* ‘looks good’). The part of speech of *feikki*, having been classified both as a noun and an adjective in the dictionary, remains ambiguous in these specific contexts.

¹¹ In this study, the term collocation is preferred to compound, as it covers both non-segmentable and segmentable collocations as well as various writing patterns. Furthermore, certain defining properties of Finnish compounds, such as stress patterns, cannot be deduced with certainty from the written data (for characteristics of prototypical compounds in Finnish, cf. Hyvärinen 2019: 310).

- (14) [segmentable; base: adjective]
kvasifilosofinen *kehäpäätelmä*
 quasi.philosophical[NOM] circular.reasoning[NOM]
 ‘quasi-philosophical circular reasoning’ (fiTenTen24)

To distinguish collocational from attributive adjectival use, the main criterion in this study is the lack of agreement in case and number in collocational constructions. In general, attributive adjectives agree with the modified noun in Finnish (Hakulinen et al. 2004: 1243). Thus, *feikki* functions as an adjective in the expression *feikkien uutisten* ‘of fake news’, showing agreement in genitive plural with the noun *uutinen* ‘news’ (15), whereas *feikkiuutisten* ‘of fake news’ is considered a collocational form, characterized by the invariant form *feikki* (16). The degree of bonding in writing may, however, vary in collocations, which means that the endless nominative singular is ambiguous between collocational and adjectival interpretation (17).

- (15) [adjectival]
feikkien uutisten ja virheellisen tiedon estämiseksi
 fake:GEN;PL news:GEN;PL and incorrect:GEN information:GEN prevent:NMZ:TRANS
 ‘in order to prevent **fake news** and misinformation’ (fiTenTen24)
- (16) [collocational]
Juttu noudattelee menestyneimpien feikkiuutisten kaavaa
 thing[NOM] comply:3SG successful:SUP:GEN;PL fake.news:GEN;PL pattern:PART
 ‘The story follows the formula of the most successful **fake news**’ (fiTenTen24)
- (17) [ambiguous]
Tuo skeittilauta juttu on kuulemma feikki uutinen.
 that[NOM] skateboard[NOM] thing[NOM] be.3SG supposedly fake[NOM] news[NOM]
 ‘That skateboard thing is supposedly **fake news**.’ (fiTenTen24)

2.3 The morphological status of *feikki*, *kvasi-* and *pseudo-*

This study uses the term ‘morpheme’ to describe *feikki*, *kvasi-* and *pseudo-*, as they form meaningful units that cannot be further segmented into smaller parts. However, more fine-grained morphological classifications have been proposed for these morphemes in the literature on other languages: Cappelle et al. (2023) use the term ‘prefix’ for *quasi-* and *pseudo-* in English, focusing on their initial position; in their paper on *pseudo-* in Greek and French, Vassiliadou et al. (2023: 237) list different existing classification proposals from the relevant literature, ranging from “prefix or prefixoid” to “bound morpheme entering neo-classical compositions” and “semiword”, while Van Goethem et al. (2025: 27) refer to *pseudo-* as “a ‘neo-classical combining form’, which reflects its special status in morphology as a bound morpheme of Ancient Greek origin.” In her study of combining forms in English, Mattiello (2023: 2-3) defines them “as initial or final bound morphemes which are either allomorphic variants of classical Latin or Greek words [...] or shortenings of [...] English words”. In Mattiello’s approach, they are examples of “transitional morphology, i.e., intermediate between two subcomponents of word-formation: compounding and derivation” (Mattiello 2023: 3), a characterization that is consistent with the different classification proposals given above (for a systematic comparison of properties of affixes, affixoids, combining forms and compound constituents, see Mattiello 2023: 49).

While the exact morphological status of the three morphemes, be it combining form for *pseudo-* and *kvasi-*, or free lexeme used in a more prototypical type of compounding for *feikki*, certainly calls for further study, this paper focuses on their shared use as modifying first constituents in collocations that can be considered approximative formations. Therefore, even though their status and degree of lexical autonomy may differ, *feikki*, *pseudo-* and *kvasi-* are treated together under the cover term ‘morpheme’.

3 Research questions

This study aims to answer the following research questions:

RQ1: In (segmentable) collocational use, what is the part of speech of the collocates the borrowed morphemes combine with?

RQ2: How productive are the borrowed morphemes in segmentable collocations?

RQ3: Which collexemes are attracted to the collocational construction?

4 Data and methods

Data are drawn from the fiTenTen24 web corpus at Sketchengine¹² (Kilgarrieff et al. 2014); this corpus comprises 4.4 billion words and contains a variety of genres and writing styles crawled from the Internet, including, among others, texts from blogs and discussion boards as well as fictional and encyclopaedic texts.

The regular expressions in (18)–(20) were used to find instances where the morphemes were used in collocations in bound, hyphenated and free form.

(18) [word=".*[Ff]eik.*" & lemma!="feikata"]¹³

(19) [word=".*[Kk]vasi.* "]

(20) [word=".*[Pp]pseudo.* "]

For each morpheme a sample of 3000 concordance lines, shuffled once in SketchEngine, was downloaded and manually annotated in Excel until 500 relevant occurrences were obtained. The sampling of 3000 concordance lines ensured the inclusion of occurrences distributed throughout the whole corpus, while shuffling yielded a random order within the sample. Instances in the sample that were not relevant were discarded; these include use of the morpheme in question as an inflected adjective or as a derivational base, doublets, meta-linguistic commentaries, use of the morphemes in user names and URLs and in non-Finnish co(n)texts.¹⁴ Furthermore, I also discarded occurrences in texts from a selected list of websites that contained many grammatical errors.¹⁵

¹² Available at <https://www.sketchengine.eu/>.

¹³ For omission of the verb *feikata* ‘to fake’, although, due to incorrect tagging, the verb still was frequently encountered in the raw data.

¹⁴ Although non-nativized *fake* and *quasi-* are attested in the corpus as well (albeit much less frequently), they were not included in this study, as their use seems to be restricted to mostly non-Finnish contexts (i.e., English, Italian or Latin sentences).

¹⁵ This pertains to certain website domains that offer automatic translations of Wikipedia articles (abcdef.wiki, wiki7.org, wiki34.com and wikiital.com).

The occurrences were annotated for the following properties:

- Construction type (non-segmentable collocation, segmentable simple collocation, segmentable embedded collocation)
- In segmentable collocations: the lemma of the collocate, its part of speech and the orthographic bonding between the morpheme and its collocate

Concerning **construction type**, non-segmentable collocations comprise neoclassical formations that do not contain “a synchronically recognizable and independent base” (Van Goethem et al. 2025: 33) and (a singular instance of) a blend (22). This is in contrast to segmentable collocations (as in the earlier examples (12)–(14), where *uggit* ‘Uggs’, *tiede* ‘science’ and *filosofinen* ‘philosophical’ are independent and recognizable bases and form, together with the borrowed morphemes, a segmentable simple collocation); a segmentable simple collocation may, itself, be part of a larger collocational structure (and thus form a segmentable embedded collocation, such as in example (23), where the collocation *kvasipredikaatti* ‘quasi-predicate’ is further embedded into *kvasipredikaattiilmaus* ‘quasi-predicate expression’).

- (21) [non-segmentable] *Sairaalapotilailla E. coli suhteellinen osuus virtsatieinfektion aiheuttajana on pienempi ja **pseudomonasten**, enterokokkien, stafylokokkien ja kandidan suurempi.*
 ‘In hospitalized patients, the relative proportion of *E. coli* as cause of urinary tract infection is lower and that of **pseudomonas**, enterococci, staphylococci and candida is higher.’ (fiTenTen24)
- (22) [blend] *Esiintyvätö tukiasemat satelliitteina (**pseudoliitteina**) AGPS-järjestelmässä*
 ‘Do the base stations appear as satellites (**pseudo[-satel]lites**) in the AGPS-system’ (fiTenTen24)
- (23) [segmentable embedded collocation] *Tämä tarkoittaa sitä, että transitiivilauseessa merkitykseltään spatiaalinen **kvasipredikaattiilmaus** kuvaa paikkaa (tai muuta tilaa) johon suhteutuu ensisijaisesti objektin tarkoite*
 ‘This means that in a transitive clause, a spatial **quasi-predicate expression** describes a place (or other space) that is primarily related to the referent of the object’ (fiTenTen24)

With the segmentable collocation types, the collocate was annotated for its part of speech (noun, proper noun, noun phrase, adjective, participle, adverb, verb and, in one ambiguous case, “noun or adjective” (NorA), cf. example (24)).

- (24) [NorA]
*olen **pseudopohjalainen***
 be:1SG pseudo.ostrobothnian[NOM]
 ‘I am [a] **pseudo-Ostrobothnian**’ (fiTenTen24)

Concerning **orthographic bonding**, four annotation values were used: bound (25), hyphen (26), free (27) and clipped (28). Ambiguous cases in the writing of *feikki* (where its status as adjective or part of a collocation cannot be determined, cf. §2.2 and example (17)), were considered collocations in the annotation process. As this might impact the results regarding

the use of free *feikki*, the proportion of ambiguous cases that received the value ‘free’ in the annotation process is reported explicitly in §0 on orthographic bonding. In a limited number of cases, one or more constituents were marked by additional punctuation, such as brackets and quotation marks (for the use of quotation marks as “scare-quotes”, cf. Schäfer 2025). In these cases, the additional punctuation was disregarded and the value was assigned based on the orthographic form that remained once the additional punctuation was removed (29).

- (25) [bound]
Haydn loi [...] oratorion, joka on kirkollis-maallinen
 Haydn create:PST[3SG] oratorio:GEN that be.3SG ecclesiastical-secular[NOM]
kvasiooppera
 kvasi.opera[NOM]
 ‘Haydn composed an [...] oratorio, which is an ecclesiastical-secular **quasi-opera**’
 (fiTenTen24)
- (26) [hyphen]
Monista [...] tuli pseudo-tieteen ja hysterian uhreja
 many:PL:ELAT come:PST[3SG] pseudo-science:GEN and hysteria:GEN victim:PL;PART
 ‘Many people [...] became victims of **pseudo-science** and hysteria’ (fiTenTen24)
- (27) [free]
saimme [...] rahamme takaisin näistä feikki kengistä!
 get:PST:1PL money:POSS.1PL back that:PL:ELAT fake shoe:PL:ELAT
 ‘We got back our money [that we paid] for those **fake shoes!**’ (fiTenTen24)
- (28) [clipped]
Tätä kutsutaan joskus pseudo- eli valeiskiakseksi. [fiTenTen24]
 this:PART call:PASS sometimes pseudo- or fake.sciatica:TRANSL
 ‘This is sometimes called **pseudo-** or false **sciatica.**’ (fiTenTen24)
- (29) [bound]
ja esitetään biologisia, [...] tai psykologisia
 and present.PASS biological:PART;PL or psychological:PART;PL
(pseudo)teorioita
 pseudo.theory:PART;PL
 ‘and they present biological, [...] or psychological **(pseudo)theories**’ (fiTenTen24)

The productivity of the morphemes is calculated as type/token ratio (TTR) and hapax/token ratio, or potential productivity (PP) (Baayen 2009). In order to identify significantly attracted collexemes, I used the Multiple Distinctive Collexeme Analysis outlined in Levshina (2015: 248–251) and Stefanowitsch & Gries (2003).

5 Results

5.1 Construction types, part of speech of the collocate and word frequencies

Figure 1 shows the distribution of construction types (token frequency). In all three cases, the largest proportion consists of segmentable simple collocations, with the numbers ranging from 348 for *pseudo-* to 483 for *feikki*. For all morphemes, we find a smaller number of segmentable

embedded collocations, with the numbers ranging from 17 for *feikki* to 26 for *kvasi-*. Non-segmentable collocations are only observed in the data sample for *pseudo-*.

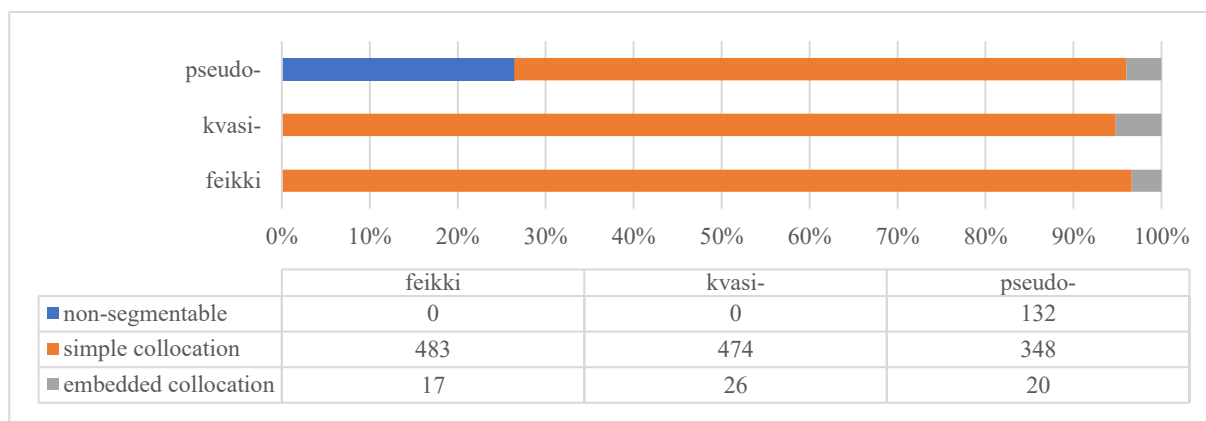


Figure 1: Construction types (based on token frequency)

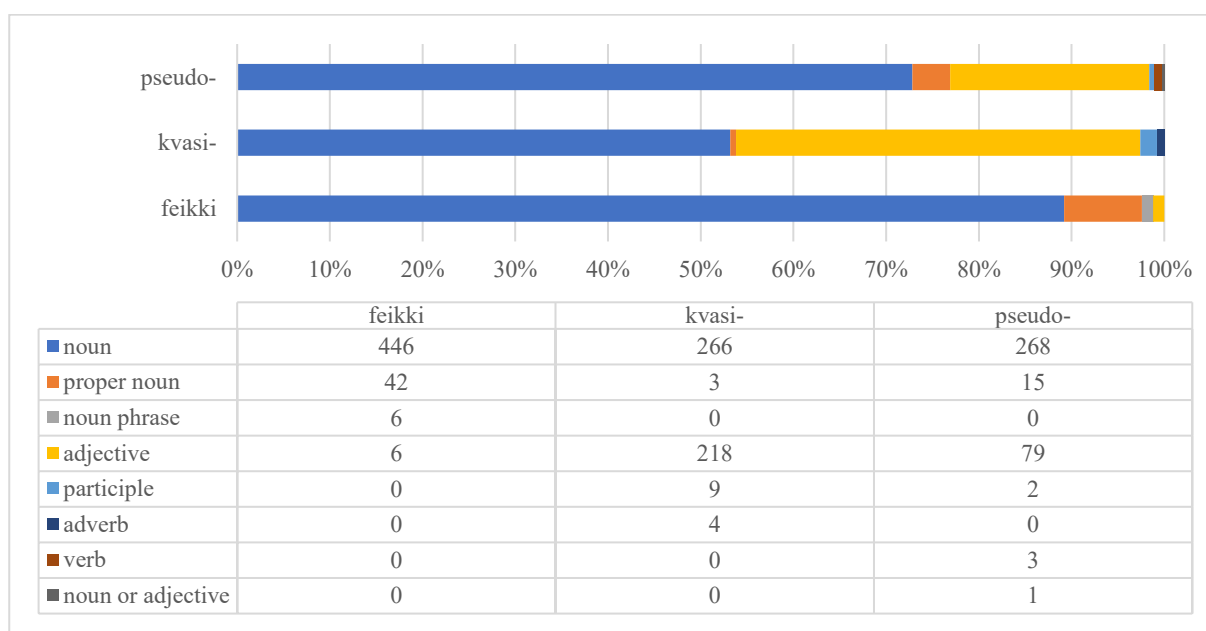


Figure 2: Part of speech of collocates (based on token frequency; excl. non-segmentable)

Figure 2 shows the part of speech of the collocate (based on token frequencies, excluding non-segmentable collocations). Nouns form the most frequent collocates for all three morphemes, accounting for more than 50% of the data within each sample. Proper nouns combine predominantly with *feikki* (e.g., *feikkiuggit* ‘fake Uggs’, *feikki-crocsit* ‘fake Crocs’, both shoe brands), to a lesser extent with *pseudo-* (e.g., *pseudo-Messias* ‘pseudo-Messiah’) and only marginally with *kvasi-* (e.g., *kvasilatina* ‘quasi-Latin’). In the case of *feikki*, combining nouns, proper nouns and noun phrases accounts for 98,8% of the data sample, indicating the highest proportion of noun collocates among the three morphemes.

Adjectival collocates, on the other hand, are most frequently encountered with *kvasi-* (e.g., *kvasitieteellinen* ‘quasiscientific’), less so with *pseudo-* (e.g., *pseudotieteellinen* ‘pseudoscientific’) and only marginally attested for *feikki* (e.g., *feikki-keskiaikainen* ‘fake medieval’). Participles, conveying both verbal and adjectival characteristics, are accordingly

found predominantly (albeit still sparsely) with *kvasi-* (e.g., *kvasi-randomoitu* ‘quasi-randomized’), only marginally with *pseudo-* (*pseudorandomoitu* ‘pseudo-randomized’) and not at all with *feikki*.

Other parts of speech, such as adverbs or verbs, only occur very rarely in the data samples and are restricted to *kvasi-* and *pseudo-*, respectively (see example (30) for an example of a verbal collocate).

- (30) *Pseudosiivotessani meidän makkaria*
 pseudo.clean:INF.POSS.1SG 1PL:GEN bedroom:PART
 ‘While I was pseudocleaning our bedroom’ (fiTenTen24)

The collocate types with a token frequency of 5 or higher (and, for *pseudo-*, also the second part in non-segmentable collocations) are shown in Table 2.

Table 2: Collocate types with a token frequency ≥ 5 ;
 colour coded: white = N or N-prop; grey = non-segmentable; orange = A

<i>feikki</i>	<i>pseudo-</i>	<i>kvasi-</i>
<i>profili</i> ‘profile’ (28)	<i>tiede</i> ‘science’ (54)	<i>ryhmä</i> ‘group’ (32)
<i>uutinen</i> ‘news’ (13)	<i>-nyymi</i> ‘pseudonym’ (41)	<i>kokeellinen</i> ‘experimental’ (30)
<i>tili</i> ‘account’ (10)	<i>tieteellinen</i> ‘scientific’ (28)	<i>staattinen</i> ‘static’ (26)
<i>kuva</i> ‘picture’ (10)	<i>-nymisoida</i> ‘pseudonymize’ (23)	<i>tieteellinen</i> ‘scientific’ (20)
<i>laukku</i> ‘bag’ (10)	<i>-monas</i> ‘pseudomonas’ (18)	<i>tiede</i> ‘science’ (19)
<i>tuote</i> ‘product’ (8)	<i>koodi</i> ‘code’ (16)	<i>säännöllinen</i> ‘regular’ (15)
<i>lasit</i> ‘glasses’ (7)	<i>-nymisointi</i> ‘pseudonymization’ (13)	<i>kide</i> ‘crystal’ (12)
<i>ommel</i> ‘seam, stitch’ (7)	<i>-tuberculosis</i> ‘pseudotuberculosis’ (12)	<i>hyperbolinen</i> ‘hyperbolic’ (9)
<i>versio</i> ‘version’ (6)	<i>manga</i> ‘manga’ (11)	<i>konformikuvaus</i> ‘conformal map’ (8)
<i>puku</i> ‘suit, clothing’ (5)	<i>tieteilijä</i> ‘scientist’ (8)	<i>predikaatti</i> ‘predicate’ (8)
<i>lippu</i> ‘ticket’ (5)	<i>efedriini</i> ‘ephedrine’ (7)	<i>looginen</i> ‘logical’ (7)
<i>video</i> ‘video’ (5)	<i>geeni</i> ‘gene’ (5)	<i>tieto</i> ‘data, information’ (7)
<i>crocs</i> ‘Crocs’ (5)		<i>filosofinen</i> ‘philosophical’ (7)
		<i>rakenne</i> ‘construction’ (7)
		<i>hiukkanen</i> ‘particle’ (6)
		<i>resultatiivinen</i> ‘resultative’ (6)
		<i>markkina</i> ‘market’ (6)
		<i>geoidi</i> ‘geoid’ (5)
		<i>realismi</i> ‘realism’ (5)
		<i>uskonnollinen</i> ‘religious’ (5)

Table 2 is revealing in several aspects. Regarding the part of speech, nouns are shown to be frequent collocate types with all three morphemes. In fact, the most frequently observed collocate type for each morpheme is a noun. Frequent adjectival collocate types, however, draw a different picture across the data sets: while a high proportion of the most frequent collocate types of *kvasi-* consists of adjectives, not a single adjective occurs as a frequent type with *feikki* (which is in line with the token frequencies as seen in Figure 2). Only one adjective (*tieteellinen* ‘scientific’) occurs as a frequent type base with *pseudo-*. This very adjective and its corresponding noun *tiede* ‘science’ form the only overlap in frequent collocate types across the table, both of them occurring with *pseudo-* and *kvasi-* (for shared collocate types with a lower token frequency than 5 in two or more of the datasets, see §0). Frequent non-segmentable collocation types feature with high token frequency in the *pseudo*-data sample (the most frequent type thereof being *pseudonnyymi* ‘pseudonym’).

Regarding the semantic profile of the most frequent collocate types, for *feikki* many of them are related either to digital media (such as *profili* ‘profile’, *uutinen* ‘news’, *tili* ‘account’, *video* ‘video’) or refer to concrete shopping items (such as *laukku* ‘bag’, *tuote* ‘product’, *lasit* ‘glasses’, *puku* ‘suit’). By contrast, many frequent collocate types of *kvasi-* and *pseudo-* relate to the domain of science and scientific terminology (e.g., the mathematical term *quasiryhmä* ‘quasigroup’, the academic term *quasikokeellinen* ‘quasi-experimental’, the medical substance *pseudoefedriini* ‘pseudoephedrine’ or the biological term *pseudogeeni* ‘pseudogene’).

5.2 Orthographic bonding

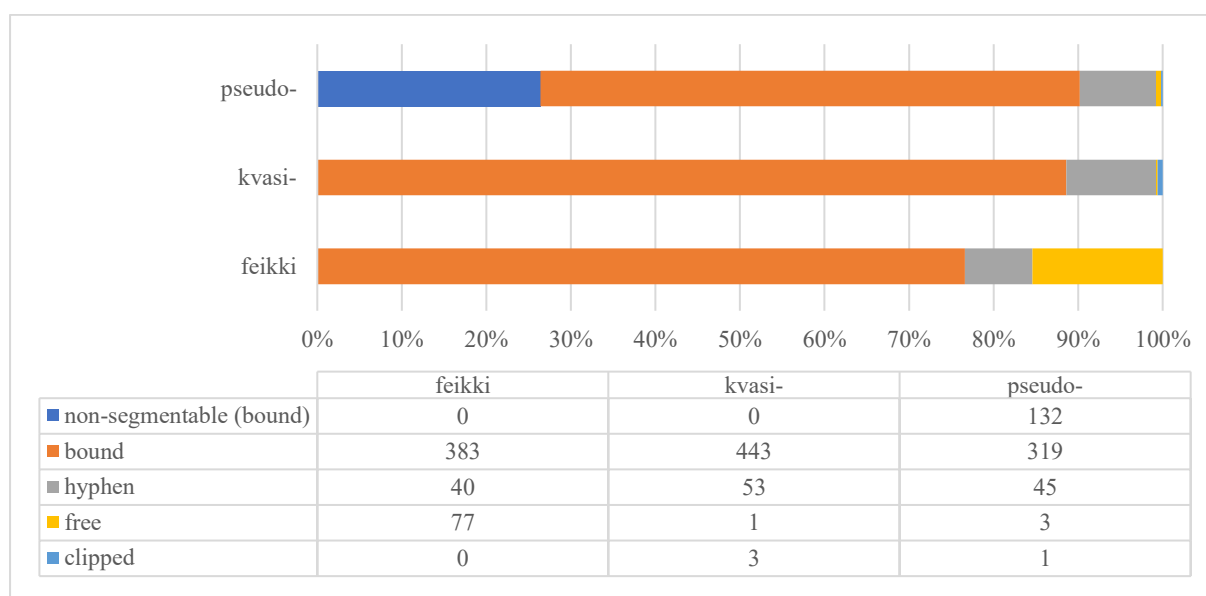


Figure 3: Orthographic bonding (based on token frequency)

Figure 3 shows the orthographic bonding (based on token frequencies). Bound use is predominant (and exclusive for non-segmentable collocations). This is not surprising, as it is the general orthographic rule compounds are subject to in Finnish (Hyvärinen 2019: 311), although borrowed morphemes may deviate from this pattern (e.g., *wannabe*; cf. Norde et al. 2025: 368–369).

Free use in collocations is found predominantly with *feikki*, while it is very rare with *kvasi-* and *pseudo-* (with only one free occurrence for *kvasi-* and only three free occurrences

for *pseudo-*). The relatively high proportion of free use with *feikki* can be accounted for, at least in part, by the ambiguity in the endingless nominative singular, where it could also be analysed as an attributive adjective (cf. §0). Of the 77 free occurrences of *feikki*, 35 (45%) are directly followed by a noun in the nominative singular and thus ambiguous. This, however, also shows that more than half of the free occurrences of *feikki* in the sample cannot be attributed to this ambiguity (because the nominal base of the collocation is not in the nominative singular; cf., e.g., example (27) above), which indicates that *feikki* is more often used as a free form than the other two morphemes, even if we do not take the ambiguous cases into account.

Hyphenated writing, on the other hand, is of comparable frequency for all three morphemes. A proportion of these instances can be explained by prescriptive Finnish orthographic rules, which call for a hyphen in specific contexts, such as consecutive adjacent vowels at the constituent boundary (as in *pseudo-okkultismi* ‘pseudo-occultism’) or when one constituent is an abbreviation (such as *feikki-OVH* ‘fake RRP [recommended retail price]’) (cf. Hakulinen et al. 2004: 389–390 or Hyvärinen 2019: 311 for a more detailed description). These orthographical conventions account for 55% (i.e., 22 out of 40) of the hyphenated occurrences of *feikki*, for 42,2% (19/45) of the hyphenated occurrences of *pseudo-* and for 15,1% (8/53) of the hyphenated occurrences of *kvasi-*.

However, the data also show instances that do not apply prescriptive rules of hyphenation (such as, e.g., *pseudoosoite* ‘pseudo-address’ < *osoite* ‘address’ or *feikki 4K* ‘fake 4K’).¹⁶ This suggests a certain degree of freedom regarding orthographic variation – especially in non-redacted texts. In the case of *pseudo-*, one of its more frequent collocate types, *manga* ‘manga’, is found in the data set written free (31), bound (32), and hyphenated (33), demonstrating free variation in orthographic bonding (cf. the following examples, that all stem from different URL sources):

- (31) *Vertasin kylläkin pseudo mangaa “normaaliin” tapaan piirtää*
 compare:PST:1SG yes:CLT pseudo manga:PART normal:ILL way:ILL draw:INF
 ‘I did compare **pseudo manga** to the “normal” way of drawing’ (fiTenTen24)
- (32) *Itse kutsuin ainakin omia töitáni pseudomangaksi*
 self call:PST:1SG at.least own:PL:PART work:PL:PART:POSS.1SG pseudo.manga:TRANS
 ‘I myself called at least my own works **pseudomanga**’ (fiTenTen24)
- (33) *Kun luin “pseudo-mangan”*
 when read:PST:1SG pseudo-manga:GEN
 ‘When I read the “**pseudo-manga**”’ (fiTenTen24)

Clipped instances are not very frequent, but all observed occurrences contrast the loan morpheme with a similar (near-synonymous) Finnish morpheme, as in e.g., *kvasi- eli tekotiede* ‘quasi- or false science’ or example (28) above.

5.3 Productivity

Figure 4 shows a comparison of the productivity of the morphemes in segmentable formations across all parts of speech, as indicated by type/token ratio (TTR) (the number of types divided

¹⁶ The example, with a slightly extended context, is *feikki 4K:n kohdalla* ‘regarding fake 4K’. The postposition *kohdalla* ‘regarding’ governs genitive case (regularly marked on *4K* with the inflectional ending *-n*); therefore, in this specific case *feikki* cannot be an attributive adjective, as it would have to agree in case with the head it modifies.

by the number of tokens) and potential productivity (PP) (the number of hapax legomena divided by the number of tokens) (Baayen 2009). *Feikki* has the highest scores as regards both TTR and PP, while *kvasi-* has the lowest scores and *pseudo-* is situated in-between.

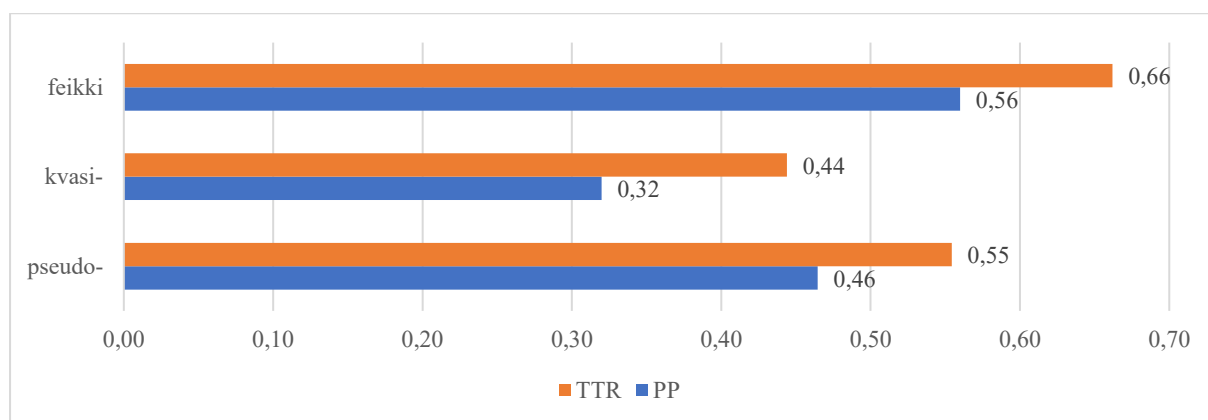


Figure 4: Potential productivity and type/token ratio (excl. non-segmentable collocations)

5.4 Shared collocates

Table 3 shows the number of types shared for each possible pairing, also taking the part of speech of the collocate into account. Across the three morphemes, the number of shared collocates is relatively small. Only one collocate type, *tutkimus* ‘study’, is represented in all three samples (*feikki* 3, *kvasi* 3, *pseudo* 1, number of tokens). As mentioned in §5.1, only two collocate types (*tiede* ‘science’ and *tieteellinen* ‘scientific’) occur frequently with more than one morpheme, both forming collocations with *pseudo-* and *kvasi-*. The overlap between *pseudo-* and *kvasi-* extends further to other collocates with lower token frequencies that are related via derivational patterns (*tieteellisyys* ‘scientificity’ and the pair *uskonto* ‘religion’ – *uskonnollinen* ‘religious’).

The first two lines of

Table 3 show that *feikki* shares more collocate types (and also a bit more overlap as regards the part of speech criterion) with *pseudo-* than with *kvasi-*. The third line indicates that *kvasi-* and *pseudo-* share both a higher number of collocates with each other, an example shown in (34)–(35), and also show more overlap regarding the parts of speech of the collocate. Considering shared collocates, this puts *pseudo-* in a central position.

It is noteworthy that both *feikki* and *kvasi-* share more collocates with *pseudo-* than with each other, as the total number of synchronically segmentable collocate tokens for the latter was, due to the exclusion of non-segmentable formations, lower than the number of segmentable collocate tokens for *feikki* and *kvasi-*.

- (34) *Brian alkaa [...] puhua pseudo-uskonnollista potaskaa*
 Brian begin:3SG talk:INF pseudo-religious:PART crap:PART
 ‘Brian starts [...] talking **pseudo-religious** crap’ (fiTenTen24)

- (35) *tuo kvasi-uskonnollinen muistopäivä [...] tulee aikaisin.*
 that[NOM] kvasi-religious[NOM] holiday[NOM] come:3SG early
 ‘that **quasi-religious** holiday [...] is coming early’ (fiTenTen24)

Table 3: Number of shared collocates (types) by part of speech of the collocate

pair	PTCP	A	N (incl. N-prop)	shared collocates
<i>feikki & kvasi-</i>	0	0	6	<i>komponentti</i> ‘component’, <i>lääkäri</i> ‘doctor’, <i>markkina</i> ‘market’, <i>porno</i> ‘porno’, <i>toimittaja</i> ‘journalist’, <i>tutkimus</i> ‘study’
<i>feikki & pseudo-</i>	0	1	10	<i>keskiaikainen</i> ‘medieval’, <i>4K</i> ‘4K (high resolution)’, <i>dokumentti</i> ‘documentary’, <i>juttu</i> ‘thing’, <i>kieli</i> ‘language’, <i>koodi</i> ‘code’, <i>nimi</i> ‘name’, <i>persona</i> ‘person’, <i>tunnus</i> ‘username’, <i>tutkimus</i> ‘study’, <i>versio</i> ‘version’
<i>kvasi- & pseudo-</i>	1	11	10	<i>randomoitu</i> ‘randomized’, <i>älykäs</i> ‘smart’, <i>filosofinen</i> ‘philosophical’, <i>historiallinen</i> ‘historic’, <i>juridinen</i> ‘juridical’, <i>looginen</i> ‘logical’, <i>rationaalinen</i> ‘rational’, <i>realistinen</i> ‘realistic’, <i>satunnainen</i> ‘random’, <i>syvälinen</i> ‘deep’, <i>tieteellinen</i> ‘scientific’, <i>uskonnollinen</i> ‘religious’, <i>latina</i> ‘Latin’, <i>liberaali</i> ‘liberal’, <i>pallo</i> ‘ball’, <i>rakenne</i> ‘construction’, <i>teologia</i> ‘theology’, <i>teoria</i> ‘theory’, <i>tiede</i> ‘science’, <i>tieteellisyys</i> ‘scientificity’, <i>tutkimus</i> ‘study’, <i>uskonto</i> ‘religion’

5.5 Multiple distinctive collexeme analysis

In order to determine which collocates are either attracted to or repelled from one of the constructions containing the three morphemes, a Multiple Distinctive Collexeme analysis was performed in R (following Levshina 2015: 248–251; cf. also Stefanowitsch & Gries 2003), taking $\log p \geq 1.3$ as a cut-off point for statistical attraction to the construction (it approximately corresponds to the p -value of 0.05, see Levshina 2015: 247) and $\log p \leq -1.3$ as cut-off point for repelled collexemes.

This method provided a total of 26 distinctive collexemes for *feikki*, 13 for *pseudo-* and 25 for *kvasi-*. Table 4 lists the top 10 distinctive collexemes for the three morphemes (the full list is provided in the appendix). A higher $\log p$ value can be interpreted as a stronger attraction. The list of distinctive collexemes is reminiscent of Table 2, in which the collocate types with a token frequency ≥ 5 were listed. This, however, does not come as a surprise, as the number of shared collocates is, as discussed in §5.4, relatively small. The lack of attraction of the collexemes *tiede* ‘science’ and *tieteellinen* ‘scientific’ in the *kvasi-* column is more noticeable.

Table 4: Top 10 attracted collexemes ($\log p \geq 1.3$)

	<i>feikki</i>	logp	<i>pseudo-</i>	logp	<i>kvasi-</i>	logp
1	<i>profili</i> ‘profile’	12.4529530	<i>tiede</i> ‘science’	17.1286779	<i>ryhmä</i> ‘group’	14.2691923

2	<i>uutinen</i> 'news'	5.7259927	<i>koodi</i> 'code'	8.1208479	<i>kokeellinen</i> 'experimental'	13.3598755
3	<i>kuva</i> 'picture'	4.3961782	<i>manga</i> 'manga'	6.3206500	<i>staattinen</i> 'static'	11.5484112
4	<i>laukku</i> 'bag'	4.3961782	<i>tieteellinen</i> 'scientific'	5.4974688	<i>säännöllinen</i> 'regular'	6.6153854
5	<i>tili</i> 'account'	4.3961782	<i>tieteilijä</i> 'scientist'	4.5862718	<i>kide</i> 'crystal'	5.2821528
6	<i>tuote</i> 'product'	3.5124664	<i>efedriini</i> 'ephedrine'	4.0099209	<i>hyperbolinen</i> 'hyperbolic'	3.9540402
7	<i>lasit</i> 'glasses'	3.0714552	<i>geeni</i> 'gene'	2.8598631	<i>konformikuvaus</i> 'conformal map'	3.5124664
8	<i>ommel</i> 'seam, stitch'	3.0714552	<i>skeptikko</i> 'sceptic'	2.2861501	<i>predikaatti</i> 'predicate'	3.5124664
9	<i>crocs</i> 'Crocs'	2.1911144	<i>kysta</i> 'cyst'	1.7133103	<i>tieto</i> 'information'	3.0714552
10	<i>lippu</i> 'ticket'	2.1911144	<i>sana</i> 'word'	1.7133103	<i>hiukkanen</i> 'particle'	2.6310050

Table 5 lists the top 10 repelled collexemes (with a $\log p \leq -1,3$) for the three morphemes. In total, 16 repelled collexemes were found for *feikki*; only 7 collexemes are repelled from *pseudo*- and 9 from *kvasi*-. In contrast to the table of attracted (distinctive) collexemes, there is overlap between pairs (such as *ryhmä*, which is repelled both from *feikki* and *pseudo*-, or *profiili*, which is repelled both from *pseudo*- and *kvasi*-).

Table 5: Top 10 repelled collexemes ($\log p \leq -1,3$)

	<i>feikki</i>	logp	<i>pseudo</i> -	logp	<i>kvasi</i> -	logp
1	<i>tiede</i> 'science'	-14.8428705	<i>ryhmä</i> 'group'	-4.2286339	<i>profiili</i> 'profile'	-5.4835158
2	<i>tieteellinen</i> 'scientific'	-9.4703383	<i>kokeellinen</i> 'experimental'	-4.0048638	<i>koodi</i> 'code'	-3.2606832
3	<i>ryhmä</i> 'group'	-6.2243233	<i>profiili</i> 'profile'	-3.5379942	<i>uutinen</i> 'news'	-2.5056576
4	<i>kokeellinen</i> 'experimental'	-5.6795655	<i>staattinen</i> 'static'	-3.3419884	<i>manga</i> 'manga'	-2.0293706
5	<i>staattinen</i> 'static'	-4.9775020	<i>säännöllinen</i> 'regular'	-1.8051814	<i>kuva</i> 'picture'	-1.7758021
6	<i>säännöllinen</i> 'regular'	-2.7568769	<i>uutinen</i> 'news'	-1.5938320	<i>laukku</i> 'bag'	-1.7758021
7	<i>kide</i> 'crystal'	-2.2725087	<i>kide</i> 'crystal'	-1.3583713	<i>tili</i> 'account'	-1.7758021
8	<i>koodi</i> 'code'	-2.0515093			<i>tuote</i> 'product'	-1.5143276
9	<i>manga</i> 'manga'	-2.0293706			<i>tieteilijä</i> 'scientist'	-1.5143276
10	<i>konformikuvaus</i> 'conformal map'	-1.5143276				

6 Summary and discussion

6.1 Construction types, orthographic bonding and productivity

The results presented in the previous section indicate both differences and similarities between the use of the three morphemes concerning the segmentability of the collocations, the part of speech of the collocate they combine with, the degree of bonding and the productivity of the constructions. In many of the properties examined, *pseudo-* occupies an intermediate position between *feikki* and *kvasi-*.

Non-segmentable collocations are found only for *pseudo-*, accounting for about 26% of the tokens. Apart from one blend, all non-segmentable collocations are either borrowed internationalisms (such as *pseudonyymi* ‘pseudonym’) or form part of taxonomical nomenclature in scientific classifications (such as *Sorbus pseudofennica* ‘Hedlundia pseudofennica [plant]’).

Regarding the part of speech of the collocate, nouns form the majority for all three morphemes, with *feikki*-collocations consisting almost exclusively of nouns. *Kvasi-* and, to a lesser degree, *pseudo-* also form adjectival collocations. This pattern is also reflected in the most frequent collocate types ($n \geq 5$) and in the distinctive collexeme analysis. Thus, *pseudo-* seems to occupy a position between *feikki* and *kvasi-* regarding the part of speech of the collocate it combines with.

Concerning orthographical bonding, the bound use is predominant in all three cases. While the hyphenated use can in part be explained by (prescriptive) Finnish orthographic rules, the collocations in the data do not always comply with these rules. The free use is most frequently attested for *feikki*. The ambiguity between collocational and attributive adjectival use in the nominative singular may contribute to the higher frequency of free *feikki*, but does not fully account for it: *feikki* remains more frequent in free use than the other two morphemes even when the ambiguous cases are excluded.

Finally, regarding productivity, *pseudo-* occupies again an intermediate position between *feikki* and *kvasi-*, both in terms of type-token ratio and potential productivity.

6.2 Semantic profiles

Based on frequently occurring collocate types, the data suggest a difference in the semantic profile of *feikki*-collocates, on the one hand, and *kvasi-* and *pseudo-* collocates on the other: *feikki* combines with lexical items related to digital media or concrete shopping items, while many frequent collocate types of *kvasi-* and *pseudo-* relate to the domain of science and scientific terminology. This division is also supported by the distinctive collexeme analysis, which additionally suggests a division between *kvasi-* and *pseudo-*: while distinctive *kvasi-* collexemes concern specific terms used in different scientific fields, the generic terms *tiede* ‘science’ and *tieteellinen* ‘scientific’ are distinctive collexemes of *pseudo-* only.

While all three morphemes can be considered approximative in the literal sense of approaching a certain concept (while possibly also falling short of reaching it), the collocations they form suggest differences in the values and focus of approximation. These differences tie

in to different values of approximation mentioned in Masini et al. (2023: 11–14), among them non-authenticity, privativity (i.e., categorical exclusion), subjective depreciation and non-prototypicality. These values should not be understood as mutually exclusive, but the data suggest that the collocations containing the morphemes may foreground different values.

Although the data were not annotated for semantic categories, a qualitative survey – for illustrative purposes only – seems to support the division of labour among the morphemes further: several *feikki*-collocates are brand names (such as *feikkiaddu* ‘fake Adidas’, *feikkikola* ‘fake Cola’) or relate to lifestyle (*feikkirusetus* ‘fake tan’, *feikkitatska* ‘fake tattoo’). They suggest expression of non-authenticity and artificiality, typically conveying an active sense of forgery or attempted deception. In other words, they question the genuineness of the concept expressed by the base (e.g., *feikkiprofili* ‘fake profile’, *feikki LV* and *feikkivuittoni* ‘fake Louis Vuitton’ and *feikki-video* ‘fake video [in describing a manipulated video supposedly showing paranormal activity]’). The suggestion of non-authenticity may result in a depreciative undertone, but not necessarily: a *feikkiplugi* ‘fake ear plug’ is described positively as a great way to try out a new style before stretching the ear holes.

Pseudo-collocations suggest categorical exclusion, placing an entity outside the target category, often with a depreciative undertone: in example (3), *pseudotiede* ‘pseudo-science’ is equated with *epätiede* ‘non-science’ and in example (34) – *pseudo-uskonnollinen potaska* ‘pseudo-religious crap’ – the negative evaluation is further strengthened by the noun the collocation modifies. The depreciative undertone, however, is generally lacking in neoclassical formations (*pseudonyymi* ‘pseudonym’) and taxonomical classification (cf. Bauer et al. 2013: 416 on English *pseudo*-). A subset of *pseudo*-collocates refers to cultural genres and media forms (such as, e.g., *pseudokantri* ‘pseudo-Country [music]’, *pseudodokumentti* ‘pseudo-documentary/mockumentary’ and the distinctive collexeme *manga* ‘manga’).

Kvasi-collocations suggest partial membership or non-prototypicality within a category, especially in scientific terminology: they resemble the (simple) collocate in relevant respects, but do not fully instantiate the category or only do so in a restricted sense. An example is (2), where *kvasisatunnaistetut tutkimukset* ‘quasi-randomized studies’ are contrasted with *satunnaistetut* ‘randomized’ ones; while quasi-randomized studies are not truly random, they still form part of the scientific discourse and are accepted as studies. Apart from scientific contexts, a subset of collocates relates to ideologies and politics (such as *kvasi-demokratia* ‘quasi-democracy’, *kvasikansallissosialistinen* ‘quasi-national socialist’, *kvasifeministinen* ‘quasi-feminist’). While *kvasi*- does not seem to encode pejorative evaluation as such, such an interpretation is possible especially when the base refers to ideological or political concepts (e.g., *kvasi-demokratia* ‘quasi-democracy’).

6.3 Comparison with studies on other languages

The tendency of *fake* to combine primarily with nouns, which is also found for Finnish *feikki*, may explain why previous studies focused primarily on noun collocations (such as, e.g., Cappelle et al. 2018; however, Van Goethem & Norde 2020: 433 mention an instance of the Dutch „fake“ morpheme *nep* to modify an adjective or verb).¹⁷

¹⁷ For English, a query in enTenTen21 ([word="to"] [lemma="fake"] [tag="VV"]) produces hits like the following: *After all, if he could hire gunmen to fake rob his store, he could hire gunmen to actually rob his store*, where *fake* is used adverbially just like *actually*. Whether an analysis as a verbal compound would also be possible in cases, still remains to be addressed.

The tendency of *pseudo-* to combine with nouns more frequently than with adjectives has been reported for a number of languages (for Greek and French, cf. Vassiliadou et al. 2023: 243–244; for English, cf. Cappelle et al. 2023: 63; for Germanic and Romance languages, cf. Van Goethem et al. 2025: 35–36). Finnish does not diverge; based on the token frequency of segmentable collocations, 77% of the collocates are nouns (including proper nouns) and 21% are adjectives (cf. Figure 2).

Kvasi- also combines, based on token frequency, more often with nouns than with adjectives (~54% noun collocates (including proper nouns) and ~44% adjective collocates; cf. Figure 2). Here, *kvasi-* contrasts with findings for the use of *quasi(-)* in English, where the ratio of bases is reported to be about 30% nouns and about 70% adjectives (Cappelle et al. 2023: 63–64).

As regards the most frequent types found for *pseudo-*, Finnish differs from the Romance and Germanic languages discussed in Van Goethem et al. (2025: 36–37). In their study, the lexeme *pseudonym* is the most frequent type across all eight languages the study covers, followed by the types *science* or *scientific*, that are, however, far less frequent. In Finnish, these types are reversed (and the gap between type frequencies is not as pronounced); while *pseudonyymi* also ranks high as the second most frequent type containing *pseudo-*, the type *pseudotiede* ‘pseudoscience’ is even more frequent. This might be due to the existence of another competing synonymous lexeme in Finnish, *salanimi* ‘pseudonym’ (lit. ‘secret name’).

7 Conclusion and outlook

This study examined the use of three borrowed morphemes in Finnish (*feikki*, *kvasi-* and *pseudo-*) with a focus on collocational constructions. Regarding RQ1, all three morphemes combine predominantly with nouns, while *kvasi-* and, to a lesser extent, *pseudo-* also form adjectival collocations. In regard to RQ2, *feikki* shows the highest productivity, *kvasi-* the lowest, and *pseudo-* occupies an intermediate position. Concerning RQ3, semantic profiles of the collocates differ: *feikki*-collocates refer to shopping items and digital media, while many *kvasi-* and *pseudo-*-collocates relate to science.

The differences between the three morphemes may in part reflect differences in register, given the informal nature of *feikki* and the use of *kvasi-* and *pseudo-* in academic contexts. Future research could investigate these register differences more systematically. Furthermore, the relation of the three morphemes with other (non-borrowed) morphological markers of approximation and fakeness (such as *näennäis-* ‘ostensible’, *teko-* ‘artificial’ or *vale-* ‘false’) could prove particularly enlightening.

Acknowledgements

I thank two anonymous reviewers, Muriel Norde, the editors of the volume, and Elsi Kaiser for their feedback and helpful suggestions, which greatly improved the quality of this paper.

Abbreviations¹⁸

¹⁸ Listed are abbreviations that are not part of the list of standard abbreviations in the Leipzig Glossing Rules (Comrie et al. 2015).

A adjective

CLT clitic

N noun

N-prop proper noun

NorA noun or adjective

PP potential productivity

TTR type/token ratio

V verb

Appendix: Full results of the Distinctive Collexeme analyses ($\log p \geq 1.3$ and $\log p \leq -1.3$)

	attracted					repelled				
	R1	feikki	pseudo	kvasi	logpCE	R1	feikki	pseudo	kvasi	logpCE
results for <i>feikki</i>	<i>profili</i> 'profile'	28	0	0	12.4529530	<i>iede</i> 'science'	0	54	19	-14.8428705
	<i>uutinen</i> 'news'	13	0	0	5.7259927	<i>tieteellinen</i> 'scientific'	0	28	20	-9.4703383
	<i>kuva</i> 'picture'	10	0	0	4.3961782	<i>ryhmä</i> 'group'	0	0	32	-6.2243233
	<i>laukku</i> 'bag'	10	0	0	4.3961782	<i>kokeellinen</i> 'experimental'	0	0	30	-5.6795655
	<i>tili</i> 'account'	10	0	0	4.3961782	<i>staattinen</i> 'static'	0	0	26	-4.9775020
	<i>tuote</i> 'product'	8	0	0	3.5124664	<i>säännöllinen</i> 'regular'	0	0	15	-2.7568769
	<i>lasit</i> 'glasses'	7	0	0	3.0714552	<i>kide</i> 'crystal'	0	0	12	-2.2725087
	<i>ommel</i> 'seam, stitch'	7	0	0	3.0714552	<i>koodi</i> 'code'	1	16	0	-2.0515093
	<i>crocs</i> 'Crocs'	5	0	0	2.1911144	<i>manga</i> 'manga'	0	11	0	-2.0293706
	<i>lippu</i> 'ticket'	5	0	0	2.1911144	<i>tieteilijä</i> 'scientist'	0	8	0	-1.5143276
	<i>puku</i> 'suit, clothing'	5	0	0	2.1911144	<i>konformikuvaus</i> 'conformal map'	0	0	8	-1.5143276
	<i>video</i> 'video'	5	0	0	2.1911144	<i>looginen</i> 'logical'	0	1	7	-1.5143276
	<i>versio</i> 'version'	6	1	0	1.9476168	<i>predikaatti</i> 'predicate'	0	0	8	-1.5143276
	<i>pari</i> 'pair, couple'	4	0	0	1.7517817	<i>rakenne</i> 'construction'	0	1	7	-1.5127336
	<i>rusketus</i> 'tan'	4	0	0	1.7517817	<i>filosofinen</i> 'philosophical'	0	2	7	-1.5127336
	<i>shomanssi</i> 'showmance'	4	0	0	1.7517817	<i>hyperbolinen</i> 'hyperbolic'	0	0	9	-1.5127336
	<i>sivusto</i> 'website'	4	0	0	1.7517817					
	<i>tyyppi</i> 'type, guy'	4	0	0	1.7517817					
	<i>uggit</i> 'Uggs'	4	0	0	1.7517817					
	<i>blogi</i> 'blog'	3	0	0	1.3130054					
	<i>hymy</i> 'smile'	3	0	0	1.3130054					
	<i>kilpi</i> 'license plate'	3	0	0	1.3130054					
	<i>liha</i> 'meat'	3	0	0	1.3130054					
	<i>sivu</i> 'page'	3	0	0	1.3130054					
	<i>timantti</i> 'diamond'	3	0	0	1.3130054					
	<i>traileri</i> 'trailer'	3	0	0	1.3130054					
results for <i>pseudo</i>	R1	feikki	pseudo	kvasi	logpCE	R1	feikki	pseudo	kvasi	logpCE
	<i>iede</i> 'science'	0	54	19	17.1286779	<i>ryhmä</i> 'group'	0	0	32	-4.2286339
	<i>koodi</i> 'code'	1	16	0	8.1208479	<i>kokeellinen</i> 'experimental'	0	0	30	-4.0048638
	<i>manga</i> 'manga'	0	11	0	6.3206500	<i>profili</i> 'profile'	28	0	0	-3.5379942
	<i>tieteellinen</i> 'scientific'	0	28	20	5.4974688	<i>staattinen</i> 'static'	0	0	26	-3.3419884
	<i>tieteilijä</i> 'scientist'	0	8	0	4.5862718	<i>säännöllinen</i> 'regular'	0	0	15	-1.8051814
	<i>efedriini</i> 'ephedrine'	0	7	0	4.0099209	<i>uutinen</i> 'news'	13	0	0	-1.5938320
	<i>geeni</i> 'gene'	0	5	0	2.8598631	<i>kide</i> 'crystal'	0	0	12	-1.3583713
	<i>skeptikko</i> 'sceptic'	0	4	0	2.2861501					
	<i>kysta</i> 'cyst'	0	3	0	1.7133103					
	<i>sana</i> 'word'	0	3	0	1.7133103					
	<i>vilja</i> 'cereal'	0	3	0	1.7133103					
	<i>satunnainen</i> 'random'	0	4	1	1.6914757					
	<i>kieli</i> 'language'	1	4	0	1.6914757					
results for <i>kvasi</i>	R1	feikki	pseudo	kvasi	logpCE	R1	feikki	pseudo	kvasi	logpCE
	<i>ryhmä</i> 'group'	0	0	32	14.2691923	<i>profili</i> 'profile'	28	0	0	-5.4835158
	<i>kokeellinen</i> 'experimental'	0	0	30	13.3598755	<i>koodi</i> 'code'	1	16	0	-3.2606832
	<i>staattinen</i> 'static'	0	0	26	11.5484112	<i>uutinen</i> 'news'	13	0	0	-2.5056576
	<i>säännöllinen</i> 'regular'	0	0	15	6.6153854	<i>manga</i> 'manga'	0	11	0	-2.0293706
	<i>kide</i> 'crystal'	0	0	12	5.2821528	<i>kuva</i> 'picture'	10	0	0	-1.7758021
	<i>hyperbolinen</i> 'hyperbolic'	0	0	9	3.9540402	<i>laukku</i> 'bag'	10	0	0	-1.7758021
	<i>konformikuvaus</i> 'conformal map'	0	0	8	3.5124664	<i>tili</i> 'account'	10	0	0	-1.7758021
	<i>predikaatti</i> 'predicate'	0	0	8	3.5124664	<i>tuote</i> 'product'	8	0	0	-1.5143276
	<i>tieto</i> 'data, information'	0	0	7	3.0714552	<i>tieteilijä</i> 'scientist'	0	8	0	-1.5143276
	<i>hiukkanen</i> 'particle'	0	0	6	2.6310050					
	<i>resultatiivinen</i> 'resultative (A)'	0	0	6	2.6310050					
	<i>looginen</i> 'logical'	0	1	7	2.3339155					
	<i>rakenne</i> 'construction'	0	1	7	2.3339155					
	<i>geoidi</i> 'geoid'	0	0	5	2.1911144					
	<i>realismi</i> 'realism'	0	0	5	2.1911144					
	<i>filosofinen</i> 'philosophical'	0	2	7	1.8448941					
	<i>kuu</i> 'moon'	0	0	4	1.7517817					
	<i>resultatiivi</i> 'resultative (N)'	0	0	4	1.7517817					
	<i>satunnaistettu</i> 'randomized'	0	0	4	1.7517817					
	<i>valtio</i> 'state'	0	0	4	1.7517817					
	<i>klassinen</i> 'classical'	0	0	4	1.3130054					
	<i>konvekssi</i> 'convex'	0	0	3	1.3130054					
	<i>stellaarinen</i> 'stellar'	0	0	3	1.3130054					
	<i>vahvasti</i> 'strongly'	0	0	3	1.3130054					
	<i>tähti</i> 'star'	0	0	3	1.3130054					

References

- Baayen, R. Harald. 2009. Corpus linguistics in morphology: Morphological productivity. In Lüdeling, Anke & Kytö, Merja (eds.), *Corpus linguistic: An international handbook*, 899–919. Berlin: De Gruyter Mouton. doi:10.1515/9783110213881.2.899.
- Battarbee, Keith. 2004. Finnish. In Görlach, Manfred (ed.), *English in Europe*, 261–276. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199273102.001.0001.
- Bauer, Laurie. 2006. Competition in English Word Formation. In Kemenade, Ans van & Los, Bettelou (eds.), *The Handbook of the History of English*, 177–198. Malden: Blackwell. doi:10.1002/9780470757048.ch8.
- Bauer, Laurie & Lieber, Rochelle & Plag, Ingo. 2013. *The Oxford Reference Guide to English Morphology*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198747062.003.0018.
- Cappelle, Bert & Daus, Robert & Hartmann, Stefan. 2023. The English privative prefixes *near-*, *pseudo-* and *quasi-*: Approximation and ‘disproximation’. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 52–75. doi:10.21248/zwjw.2023.1.35.
- Cappelle, Bert & Denis, Pascal & Keller, Mikaela. 2018. Facing the facts of *fake*: A distributional semantics and corpus annotation approach. *Yearbook of the German Cognitive Linguistics Association* 6(1). 9–42. doi:10.1515/gcla-2018-0002.
- Comrie, Bernard & Haspelmath, Martin & Bickel, Balthasar. 2015. The Leipzig Glossing Rules: Conventions for interlinear morpheme-by-morpheme translation. Max Planck Institute for Evolutionary Anthropology and University of Leipzig.
- Finkbeiner, Rita & Meibauer, Jörg & Wiese, Heike (eds). 2016. *Pejoration*. Amsterdam: John Benjamins. doi:10.1075/la.228.
- Grandi, Nicola & Körtvélyessy, Livia. 2015. Introduction: Why Evaluative Morphology? In Grandi, Nicola & Körtvélyessy, Livia (eds.), *Edinburgh Handbook of Evaluative Morphology*, 3–20. Edinburgh: Edinburgh University Press. doi:10.1515/9780748681754.
- Hakulinen, Auli & Vilkuna, Maria & Korhonen, Riitta & Koivisto, Vesa & Heinonen, Tarja Riitta & Alho, Irja. 2004. *Iso suomen kielioppi. [The big Finnish grammar.]* Helsinki: Suomalaisen Kirjallisuuden Seura.
- Hyvärinen, Irma. 2019. Compounds and multi-word expressions in Finnish. In Schlücker, Barbara (ed.), *Complex Lexical Units*, 307–335. Berlin: De Gruyter. doi:10.1515/9783110632446-011.
- Karlsson, Fred. 2017. *Finnish: A Comprehensive Grammar*. New York: Routledge. doi:10.4324/9781315743547.
- Kilgariff, Adam & Baisa, Vít & Bušta, Jan & Jakubíček, Miloš & Kovář, Vojtěch & Michelfeit, Jan & Rychlý, Pavel & Suchomel, Vít. 2014. The Sketch Engine: ten years on. *Lexicography* 1(1). 7–36. doi:10.1007/s40607-014-0009-9.
- Levshina, Natalia. 2015. *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam: John Benjamins. doi:10.1075/z.195.

- Masini, Francesca & Norde, Muriel & Van Goethem, Kristel. 2023. Approximation in morphology: A state of the art. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 1–26. doi: 10.21248/zwjw.2023.1.100.
- Mattiello, Elisa. 2022. *Transitional morphology: Combining forms in modern English*. Cambridge: Cambridge University Press. doi:10.1017/9781009168274.
- Norde, Muriel & Masini, Francesca & Van Goethem, Kristel & Ebner, Daniel. 2025. *Wannabe* Approximatives: Creativity, routinization or both? In Arndt-Lappe, Sabine & Filatkina, Natalia (eds.), *Dynamics at the Lexicon-Syntax Interface*, 353–384. Berlin: De Gruyter. doi:10.1515/9783111321905-013.
- Schäfer, Martin. 2025. Scare Quotes as Markers of Creative Word-Formation: A Look at English *-ness* and *-ity* Hapaxes. In Arndt-Lappe, Sabine & Filatkina, Natalia (eds.), *Dynamics at the Lexicon-Syntax Interface*, 385–412. Berlin: De Gruyter. doi:10.1515/9783111321905-014.
- Schlücker, Barbara. 2026. *Like* – a morphological evaluative loan marker in German? *Folia Linguistica* 60(2). 385–412. doi:10.1515/flin-2025-0125.
- Stefanowitsch, Anatol & Gries, Stefan Th. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243. doi:10.1075/ijcl.8.2.03ste.
- Suomi, Kari & Toivanen, Juhani & Ylitalo, Riikka. 2008. *Finnish sound structure: phonetics, phonology, phonotactics and prosody* (Studia Humaniora Ouluensia 9). Oulu: Oulu University Press.
- Van Goethem, Kristel & Norde, Muriel. 2020. Extravagant “fake” morphemes in Dutch: Morphological productivity, semantic profiles and categorical flexibility. *Corpus Linguistics and Linguistic Theory* 16(3). 425–458. doi:10.1515/cllt-2020-0024.
- Van Goethem, Kristel & Norde, Muriel & Masini, Francesca. 2025. The fate of ‘pseudo-’ words: A contrastive corpus-based analysis. *Languages in Contrast* 25(1). 23–50. doi:10.1075/lic.22003.van.
- Vassiliadou, Hélène & Gerhard-Krait, Francine & Fotiadou, Georgia & Lammert, Marie. 2023. *Pseudo(-)* in French and Greek: Categorization and approximation. *Zeitschrift für Wortbildung / Journal of Word Formation* 7(1). 234–262. doi:10.21248/zwjw.2023.1.34.

Daniel Ebner
Humboldt-Universität zu Berlin
Nordeuropa-Institut
Unter den Linden 6
10099 Berlin
Germany
e-mail: daniel.ebner@hu-berlin.de