



This is a repository copy of *Explaining uncertainty and defectivity of inflectional paradigms*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/189947/>

Version: Published Version

Article:

Nikolaev, A. and Bermel, N. orcid.org/0000-0002-1663-9322 (2022) Explaining uncertainty and defectivity of inflectional paradigms. *Cognitive Linguistics*, 33 (3). pp. 585-621. ISSN 0936-5907

<https://doi.org/10.1515/cog-2021-0041>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Alexandre Nikolaev* and Neil Bermel

Explaining uncertainty and defectivity of inflectional paradigms

<https://doi.org/10.1515/cog-2021-0041>

Received April 1, 2021; accepted July 10, 2022; published online August 11, 2022

Abstract: The current study investigates how native speakers of a morphologically complex language (Finnish) handle uncertainty related to linguistic forms that have gaps in their inflectional paradigms. We analyze their strategies of dealing with paradigmatic defectivity and how these strategies are motivated by subjective contemporaneity, frequency, acceptability, and other lexical and structural characteristics of words. We administered a verb production (inflection) task with Finnish native speakers using verbs from a small non-productive inflectional type that has many paradigmatic gaps and asked participants to inflect the verbs in a given context. Inflectional uncertainty was measured by the number of different forms the participants produced for each verb. We classified produced forms that were not expected as either synonymous or novel and measured their optimal string alignment distance to expected forms. Our analyses revealed that a usage-based approach to paradigmatic defectivity fits better with the obtained results than a classical approach typically met in dictionaries and descriptive grammars. Thus, we argue, that paradigmatic defectivity can be better described as a dynamic rather than a static system, where gaps represent a continuum of possible inflectional choices rather than a lack of an inflectional variant.

Keywords: inflectional morphology; network analysis; overabundance; paradigmatic defectivity; usage-based approach

1 Introduction

In languages with inflectional morphology, language users can experience difficulties while inflecting certain words (e.g., producing the English past tense or participle forms of verbs such as *output*). Often this expression of uncertainty can

*Corresponding author: Alexandre Nikolaev, General Linguistics, University of Eastern Finland, P.O. Box 4, Joensuu 80101, Finland, E-mail: alexandre.nikolaev@uef.fi. <https://orcid.org/0000-0001-8634-5947>

Neil Bermel, School of Languages and Cultures, University of Sheffield, Sheffield, South Yorkshire, UK, E-mail: n.bermel@sheffield.ac.uk

result in the perception that there is a gap in the word's inflectional paradigm. This is then formalized as a description of an underlying condition in which some inflected word form of a given lexeme is not in use, or there is no largely accepted candidate to fill a certain paradigmatic slot.

This condition is often posited to be the result of various morphophonological changes that have taken place in diachrony, or of semantic motivations (e.g., lack of singular forms in *pluralia tantum* nouns such as *clothes*). A common starting point is that gaps often arise in cells where more than one appropriate form could be envisaged based on different models or analogies, but as Sims (2015: 249) shows, treating defective slots as epiphenomena resulting from a clash between independent generalizations is not a universally satisfactory explanation. As we will show, in certain kinds of tasks, such a clash more frequently results in a produced form rather than an absent one. Such explanations can thus fall short when confronted with the variety of ways defectivity is expressed in usage, and, building on previous research in the field, we demonstrate using survey data that a 'gap' is not so much the absence of any possible form, as the absence of consensus in a speech community regarding the most appropriate forms to occupy the cell.

For example, in English, compounds that have an irregular verb as their second constituent (e.g., *troubleshoot*) typically yield a defective paradigm, supposedly because the morphological properties of the second stem demand a 'strong' past tense allomorph, while the semantic connection of the compound with this stem is tenuous enough to allow it to be analyzed as a separate formation, in which that pattern would not be relevant. This example seems straightforward; however, more troublingly, in Russian, some second conjugation verbs (e.g., *дезуть* 'be rude', Pertsova 2016: 8) also tend to be defective in the 1.sg. This pattern is less predictable in comparison to English compounds with irregular verbs, so it is unclear why this clash is unresolvable only for some members of the class.¹ As a consequence of this fact, only some of the novel Russian verbs joining the second conjugation will be defective, and evidence of their defectivity may show up in the general low frequency of the 1.sg form, as well as in the appearance of a morphologically unexpected and unmotivated form alongside the expected one. Likewise, Veá and Johansson (2020) found that, in Norwegian and Swedish, defectivity of many adjectives that lack neuter singular forms is productive and it extends to cover some novel adjectives

¹ Baerman (2008: 83) points out that historically, some verbs now considered defective had attested 1.sg forms, and that there are rarely any well-founded semantic reasons for defectivity in this class; likewise, the consonant mutation that occurs between the 1.sg and other conjugated forms is otherwise so exceptionless in this paradigm as to discount any 'uncertainty' around it as a possible explanation for defective behavior.

(see also Sims-Williams and Enger 2021, who describe this defectivity as a consequence of the low frequency of use and semantic distribution of adjectives; for other examples of development of paradigmatic defectivity, e.g., in Spanish verbs, see Albright 2003, or in Greek nouns, Sims 2015, that can be explained by the frequency of use and morphophonological alternations).

1.1 ‘Switch’ versus ‘dimmer’ models of paradigmatic defectivity

We note that whatever the level of predictability of paradigmatic gaps, many descriptions of defectivity assume a SWITCH-like operation (off/on). Switch models are typically associated with descriptive grammars or dictionaries. For the sake of clarity and usability, handbooks tend to mark items as missing/not in use, or will propose theoretical forms that could be used in instances where a form tends to be avoided. The switch model proposes that words constitute a set of full paradigms with a few exceptions for defective paradigms. However, it is often unclear on what grounds an authority can decide which word has a defective paradigm and which does not. Authors of dictionaries and grammars can use language corpora for this purpose, but then many paradigmatic gaps could be accidental (see, e.g., Nikolaev and Bermel, *under review*). They could instead use introspection and rely on their intuition as experienced language users, but, as we demonstrate in the present study, people differ in their language intuition and a word form that is not acceptable for one native speaker could be acceptable for another native speaker chosen from the same population.

An alternative approach to paradigmatic defectivity would take a usage-based approach: almost all inflectional paradigms are defective to some extent, since for many words native speakers never hear/read or produce some inflected word forms (for a recent and more radical treatment of this idea, see Janda and Tyers 2021; for less radical conclusions on how the speakers fill the cells they never encounter on the basis of the forms that they do encounter, see Blevins et al. 2017). In other words, during our lives each of us encounters only a subset of all possible word forms of given lexemes, thus leaving many paradigms, practically speaking, potentially defective. This approach seems intuitively to be well suited to describe languages with rich inflectional morphology, e.g., the Finnish language, which we used as a sandpit in the current study to explore paradigmatic defectivity, and in which each noun or each verb can have hundreds or even thousands of inflected word forms (Karlsson and Koskeniemi 1985: 211).

We call this usage-based approach a DIMMER model of paradigmatic defectivity, since many words have cells whose contents are potential rather than based on

empirical experience. However, we propose to narrow the boundaries of the dimmer model so that it will not include all possible words. Members of an inflectional paradigm that are likely to be unattested should not be automatically labeled as defective if speakers would have no uncertainty about them. Unlike accidental gaps, those that manifest as uncertainty in native speakers should be included in the dimmer model of paradigmatic defectivity. However, the boundary between unattested and not defective and unattested and defective (or attested and yet defective) forms is not clear-cut, because the notion of attestation and the notion of uncertainty both include human agents rather than a rule or an algorithm. That is why in the next subsection we introduce a diagnostic approach to paradigmatic defectivity.

1.2 Symptomatic versus diagnostic approaches to paradigmatic defectivity

By a SYMPTOMATIC APPROACH to paradigmatic defectivity we mean a description of the current distribution of forms, possibly augmented by the kind of *ex post facto* arguments that Sims (2015) warned against. This approach on its own can lead to explanatory circularity: language users avoid certain word forms in a defective paradigm (e.g., 1.SG form of the Russian verb *дерзнуть* ‘be rude’), because they do not know how to produce “correct” word forms, because the distribution of forms for *дерзнуть* does not seem to contain 1.SG forms, because language users avoid these word forms.

A DIAGNOSTIC APPROACH to paradigmatic defectivity takes place when one attempts to study underlying conditions prompting a certain sort of behavior in language users – one that eventually leads to the current distribution of forms. Hockett (1954) points out that much traditional morphological scholarship in the Item and Process camp accomplished this by recapitulating diachronic processes to arrive at synchronic word forms. However, even though, without doubt, there are some diachronic factors that have influenced the current distribution of forms (see, e.g., Baerman 2011 on how homophony avoidance has influenced paradigms of the Russian indefinite pronouns *некто* ‘somebody, a certain’ and *нечто* ‘something’), naïve language users are not guided by a knowledge of the historical development of defective lexemes when they inflect, e.g., the verb *дерзнуть* ‘be rude’ in the 1.SG form (see, e.g., Baerman 2008, Pertsova 2016, Sims 2015 for criticism of such implicit assumptions; however, for a proposal on integrating diachronic and synchronic factors in morphological processing, see Kapatsinski (2022); a discussion of how to model homophony avoidance as a language learning problem using diachronic information can be found in Yin and White (2022)). We aim to use

a diagnostic approach in the present study in order to explore how paradigmatic defectivity manifests itself in language use in a morphologically complex language. We aim to explore strategies that native adult Finnish speakers employ to overcome uncertainty while inflecting words with defective paradigms.

1.3 Strategies to overcome uncertainty

One obvious solution for a language user who experiences difficulties with production of a certain word form due to a paradigmatic gap is to avoid this particular word form, perhaps substituting a synonym or a paraphrase. Alternatively, if one decides or is explicitly asked to produce a word form that is considered to constitute a paradigmatic gap, one can use analogy and predict a word form based on other members of this word's paradigm, or borrow an inflectional schema/pattern from some other paradigm that has no gaps. Both strategies could lead either to an expected² word form (given the rest of the paradigm), or to a novel word form. At an individual level, a language user's uncertainty with a given word form results in the production of one form: expected, synonymous, periphrastic, or novel. At a collective level (with many language users involved), this collective uncertainty leads to overabundance, when two or more forms compete for the same slot in a paradigm (Thornton 2011). We depict this phenomenon in Figure 1.

Uncertainty in production reflects several cognitive processes such as word retrieval, word composition, and inhibition. The mere existence of different routes (reflected at the level of collective choices, see Figure 1) to filling a paradigmatic gap allows for serial or simultaneous activation of all/some of these routes while a speaker confronts a paradigmatic gap. A word form produced by the speaker reflects only one end result of this process of activation of possible multiple routes. However, the process itself presumably has some cognitive costs, and these costs are reflected in what we call uncertainty.

As Harmon and Kapatsinski (2017: 30) point out, word retrieval demands that are inherent to production can push the speaker to produce the most accessible form to express a meaning, even when a less accessible form would be a better fit to the meaning. However, as this study will show (see also a study by Bermel et al. *under review*, discussing linguistic behavior in Czech language speakers) in the

² We use the term 'expected' as a more neutral variant of the form 'correct', because correctness of a form that is supposed to replace a gap in a paradigm is a controversial concept. However, in Finnish it is typically easy for a trained linguist to predict a 'correct' form given the rest of the paradigm. To be more specific, we asked participants to produce past participle forms of 59 Finnish verbs, and for 57 out of 59 verbs, an electronic version of the Basic Dictionary of Finnish, CD-perussanakirja (1997) explicitly lists correct past participle forms.

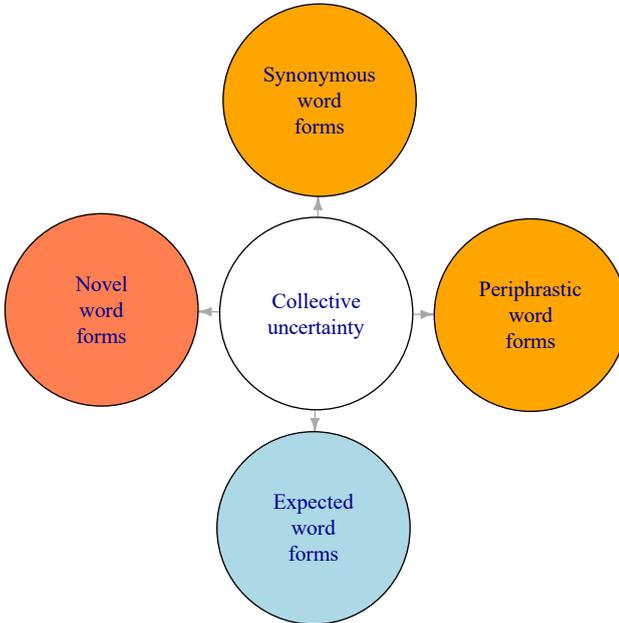


Figure 1: Overabundance as a result of defectivity. The white node (circle) in the center represents a gap in a paradigm, which language users fill with a synonymous or periphrastic (orange nodes) form, or they may produce a novel form (coral node) or an expected form (light-blue node).

case of filling a paradigmatic gap, the most accessible form will be different from person to person. Therefore, we assume that choosing the most accessible form at an individual level could reflect variation on this collective level, albeit at a smaller scale. We also propose that, because of the competition between several routes and because of inhibition due to this competition, the cognitive load of filling paradigmatic gaps should be more demanding than that of filling a regular paradigmatic slot, for which not that many options are typically available. Hence, we link uncertainty in filling paradigmatic gaps to a relatively heavier cognitive load.

1.4 Present study

We assume that paradigmatic defectivity of words (and thus collective uncertainty) will result in different inflectional choices made by native language users. However, the aim of the study is not to classify a set of Finnish verbs as more or less defective (based on how many participants chose an expected word form), but

rather to unravel structural, usage-based and cognitive factors that can explain inflectional choices/collective uncertainty related to paradigmatic defectivity. We administered one verb production task, in which participants were asked to finish a sentence by producing a past participle form for a verb presented earlier in the sentence in its past tense form. We then administered three evaluation tasks, in which other participants evaluated expected past participle forms from three different perspectives: how frequent, acceptable, or archaic/contemporary these forms are.

We mentioned earlier that we use a diagnostic approach to paradigmatic defectivity to study underlying conditions prompting a certain sort of behavior in language users. Therefore, we explain the results of the verb production (inflection) task using a set of variables collected via three subjective judgment tasks, as well as corpus frequencies of the target verbs and their lexical and grammatical characteristics. We hypothesize that all or some of these predictors contribute to uncertainty that language users demonstrate when confronted with paradigmatic gaps. We discuss these predictors in the following section in more detail.

2 Method

2.1 Materials

In choosing verbs for the present study, we started with the switch model of paradigmatic defectivity by looking in the Basic Dictionary of Finnish (1994) as well as in the Grammar of Finnish (Hakulinen et al. 2004) for verbs that have defective paradigms. After finding three verbs (**koreta*³ ‘rise’, **erata* ‘separate’, and **parata* ‘get well’) that meet the criteria for defectivity (they lack both infinitival and participle word forms in these reference works), we then loosened our criteria for the switch model of defectivity and included 56 other verbs from the same inflectional class (conjugation pattern). Since all these verbs belong to the same unproductive inflectional class, they too may lack infinitival and participle

3 We use an asterisk before the verb because as in English, in Finnish the base forms or so-called ‘dictionary’ forms of verbs are infinitive forms. For these three verbs, paradigmatic gaps occur where the infinitive forms are expected. Unlike thousands of other verbs which are listed in their infinitive forms, the two verbs can be found in an electronic version of the Basic Dictionary of Finnish, CD-perussanakirja (1997) in their third person present forms (*korkenee*, ‘he/she/it rises’, and *erkanee* ‘he/she/it separates’). The dictionary suggests that the infinitival form *parata* exists; however, the Grammar of Finnish (Hakulinen et al. 2004) disagrees with the dictionary.

word forms. Indeed, after pretesting this assumption with a few native speakers, we noticed that the three “officially” defective verbs were not that different from other verbs of this inflectional type: the other verbs also caused uncertainty (for many of these verbs, informants produced either synonymous or novel past participle forms when asked). These 56 verbs represent the dimmer model of defectivity, when a verb is not defective according to grammars or dictionaries but could be defective (at least to some extent) according to usage patterns, as reflected in language corpora or in a behavioral language experiment.

In Finnish, defective verb paradigms are formally close (sharing a string of phonemes) with some neighbor paradigms (which, however, belong to a different conjugation type). This is possible due to verb derivation, e.g., when two different derivational suffixes (and hence two different inflectional types) converged on the same meaning. According to the Grammar of Finnish (Hakulinen et al. 2004), most, but not all, of the verbs we chose for this study are originally derived from adjectives or nouns using the derivational suffix *-ne*. Many of these adjective or noun stems can also be used to form verbs using a different derivational suffix, e.g., *-ntu/nty*. Therefore, some verbs, e.g., *pienetä:pieneni* “decrease:decreased” have synonyms such as *pienentyä:pienentyi* so that verbs sharing the same meaning belong to different inflectional classes, although for some other verbs that we used in the current study, there is no derivational variant formed by the suffix *-ntu/nty*. The fact that we used so many derived verbs could affect the generalizability of our conclusions; however, we believe that studying uncertainty triggered by derived verbs with defective paradigms should also contribute to our understanding of non-derived words with defective paradigms.

2.1.1 Verb production task

The participants were given written instructions to complete a sentence presented on a computer screen by typing an appropriate word form. We used Google forms that participants completed on their own computers. In the instruction, we politely asked them not to seek any help from the internet or other people, as we would prefer that they make their own decisions/choices. They were also explicitly asked to use the same verb as in the preceding sentence. In other words, the participants read sentence #1, in which the verb was always in the past tense form (third person, singular or plural). She or he then was asked to complete sentence #2 using the same verb, except that the sentence context required the past participle form of the verb. As an example, sentence #1 with a past tense verb (underlined) is *Pekka rohkeni astua sisään* ‘Peter dared to step in’. Sentence #2, which should elicit a past participle form, is *Kalle ei _____ astua sisään* ‘Kyle didn’t _____ to step in’

(in Finnish this context requires a past participle form; in this case *rohjennut* would be the expected form).⁴ For each of the 59 verbs, we composed a pair of sentences (see Supplementary materials, pp. 1–7). The pairs of sentences were then presented to participants in a randomized order as a list of rows (one pair on each row). Participants did not have any time limit for completing the task. On average the task lasted from 15 to 25 min.

2.1.2 Verb evaluation task: subjective contemporaneity rating

Unproductive, frozen inflectional types like the one we chose for this study typically involve words that represent an older lexical stratum, so some of the word forms (e.g., infinitival forms) could have fallen out of use. Gaps in paradigms could be attributed to speakers' perceptions of these forms as old-fashioned. We therefore decided to include this variable as a covariate in our models explaining inflectional choices.

The participants were given written instructions to give contemporaneity ratings for a list of 59 expected past participles followed by example sentences (these corresponded to sentence #2 from the verb inflection task, see Section 2.1.1 and Supplementary materials, pp. 8–13). E.g., we asked how outdated or modern the following verb form is: *rohjennut* (e.g., *Kalle ei rohjennut astua sisään.*) 'dared (e.g., Kyle hasn't dared to step in.)'. One may wonder whether it would be more useful to ask about the contemporaneity rating of the lemma rather than the potentially uncertainty-inducing participial form – in other words, to ask participants to rate the infinitive form, e.g., *rohjeta* instead of presenting them with an expected participle *rohjennut*. However, in Finnish verbs, the infinitive form does not represent the concept of lemma, since it is relatively infrequent compared to other members of the paradigm (Karlsson 1983), is not the form children are likely to acquire first, nor the form that appears as a stem in most members of a verb paradigm (cf. *to dare: she dares: I dared*, and *rohjeta : hän rohkeni : minä rohkenin*).

We used Google forms that participants completed on their own computers. In the instruction, we requested that they not seek help from the internet or other people, as we value their own judgment. Each prompt consisted of a past participle and an example sentence following that contained the participle. These were presented to each participant in a randomized order. The participants were asked to choose one of the following options: *I'm not familiar with this verb; Outdated;*

⁴ Participle and infinitive forms in Finnish share the same stem (cf. participle *rohje-nnut* and infinitive *rohje-ta* to the past tense form *rohke-ni*). A novel past participle would be, e.g., *rohkennut*, which is not found in any grammar, dictionary or corpus of Finnish.

Somewhat outdated; Somewhat modern; Modern. Participants did not have any time limit for completing the task.

2.1.3 Verb evaluation task: subjective frequency

Another variable employed in the present study is perceived frequency of use. Usage-based theories of language (e.g., Bybee 2006) argue that every encounter of a word by a language user has an effect on the word's mental representation. According to Exemplar Theory, a word's representation consists of many exemplars, which are built up from individual tokens of language use and which represent the variability that exists for a given category. Harmon and Kapatsinski (2017) showed that high frequency of a form-meaning pairing leads to a strengthening of form-meaning mappings in comprehension. The authors also found that frequent forms are extended to novel meanings in production when alternatives are relatively inaccessible. This finding partially predicts that the number of synonymous forms produced to fill paradigmatic gaps would negatively correlate with lemma frequency of the words with paradigmatic gaps and would positively correlate with frequency of these synonymous words.

Although frequency of a word is a different concept than contemporaneousness of a word, these concepts are related to each other. Words that are perceived as archaic tend not to be perceived as frequent and vice versa. The connection between frequency and contemporaneousness is not just a question of perception: archaic words are those which have fallen into relative disuse, and hence are in fact less frequent. However, we assume that, despite their partial (and maybe considerable) overlap, these variables could be associated somewhat differently with strategies of overcoming uncertainty while inflecting words with defective paradigms. There are various ways to quantify frequency, and we decided to rely on the following sources. The participants were asked to choose one of the following options: *I'm not familiar with this verb; Rare; Somewhat rare; Somewhat frequent; Frequent.* The procedure was the same as for the contemporaneousness rating task (see Section 2.1.2).

2.1.4 Verb evaluation task: subjective acceptability

Acceptability judgments have a long history in linguistics, and they have long been a topic of heated debates. E.g., acceptability judgments (sometimes operationalized through introspection by one linguist) have played a central role in theories of syntax (e.g., Chomsky 1957). Even though acceptability judgments are a second-order phenomenon (users reporting on their own perceptions, rather than actual behavior), there is a consistent line of research showing their connections to real-

world data and their possibilities for elucidating it, hence their inclusion in this study. Bader and Häussler (2009), Kempen and Harbusch (2008), and Divjak (2008) all confirm for a range of types of syntactic data that the relationship is an asymmetrical one: high-frequency items or constructions in a corpus predict high acceptability ratings, while low acceptability ratings predict low corpus frequency.

Bermel and Knittl (2012: 268) additionally propose that for the kind of tightly constrained choices available in morphological paradigms, frequency data can create certain entailments: (1) that corpus frequency ranking constrains acceptability judgments (an item will not be more frequent in a representative corpus than its competing variant but less acceptable than that variant), and (2) that a lower-frequency variant suggests that the competing variant will regardless be of high acceptability. However, the applicability of these points in situations where the paradigm is potentially incomplete is unknown.

An allied issue, *statistical preemption* (see, e.g., Goldberg 2011), explains why we consistently hear, e.g., *children* in contexts where *childs* would have seemed like a more logical form. This is an example in which one inflectional construction preempts another (for derivation, an example would be the word *youth* that preempts the word *youngness*, see Goldberg 2019). The probability of preemption (its frequency) is not in linear relation with the confidence of preemption; however, frequency plays a central role in the process of statistical preemption in that higher frequency increases confidence of preemption (Goldberg 2011). In the same vein, language users tend to accept a transitive use of the verb *vanish* (e.g., *I'm gonna vanish it*) more easily than a transitive use of the verb *disappear* (e.g., *I'm gonna disappear it*), because the latter has a higher frequency of use and hence increases users' confidence that they would have heard it by now if it was in transitive use (Theakston 2004; however, see also Harmon and Kapatsinski 2017: 25, who found in the Corpus of Contemporary American English, Davies 2008, that it is *disappear* and not *vanish* that is being used transitively in production; this is possibly due to the increased use of the former in the new, specific sense 'abduct', typically used of victims of political repression). Therefore, the frequency of use and productivity of an inflectional pattern are related phenomena, e.g., frequency of use is one of the major contributing factors to some words' immunity or resistance to productive inflectional patterns (for ontogenesis of the English language, see, e.g., Lieberman et al. 2007). Both frequency and productivity contribute to how acceptable inflected word forms are when perceived by language users: if a potential form is pre-empted, we would expect its acceptability to suffer.

As with contemporaneity and subjective frequency, we asked native Finnish speakers to give their acceptability ratings. We used the following 5-point Likert scale: *I'm not familiar with this verb*; *Questionable*; *Somewhat questionable*;

Somewhat acceptable; Acceptable. We intentionally avoided the term *grammatical* when asking participants to give their acceptability rating, because although widely used in such studies, it implies a theoretical commitment to “well-formedness” by the rules of a grammar (see inter alia Kempen and Harbusch 2008) that we wished to avoid, and its use might sway native speakers to reach for forms that sound to them “more correct” than what they might ordinarily use.

For each of the three rating tasks described above, we included filler items (59 expected past participles followed by example sentences; verbs were sampled from a different, non-defective verb type). Each task including filler items lasted approximately 20 min. The reason we did not include filler items in the production task was that it would have doubled the time for completion, potentially leading to a loss of focus/motivation in an uncontrolled environment (participants completed the task on their own computers), which could have outweighed the intended benefits of fillers (participants do not prime themselves into a pattern). As the results of the production task (without fillers) will show, there was a variation in inflectional choices between participants and within participants, so we assume that if there was some self-priming, its influence was weak.

2.2 Participants

For the verb production (inflection) task we recruited 50 native Finnish-speaking young adults (mean age 27.3 years, SD 5.9; 25 females). For the three subjective rating tasks (*contemporaneity* in Section 2.1.2, *frequency* in 2.1.3, and *acceptability* in 2.1.4) we recruited respectively 30 native Finnish-speaking young adults (mean age 27.5 years, SD 7.8; 17 females), 31 native Finnish-speaking young adults (mean age 26.8 years, SD 5.5; 16 females), and 31 native Finnish-speaking young adults (mean age 25.7 years, SD 5.9; 14 females). In total, 142 participants (72 females) took part in this study. All participants were university students, and they were compensated for their time with 5-euro vouchers.

2.3 Other predictors of uncertainty

2.3.1 Stem allomorphy

Stem allomorphy is a structural variable that contributes to variability and inhibits the productivity of inflectional paradigms. The greater the stem allomorphy, the more complex the paradigm, which eventually decreases its productivity. An unproductive paradigm is at risk of losing words that have lower frequency of use, and at the same time, newcomers (novel words, loan words) typically do not follow

these paradigms (see Kirjavainen et al. 2012 for Finnish stem allomorphy and its relation to productivity of verb paradigms, and Nikolaev et al. 2014, 2018, 2020 for stem allomorphy and noun paradigms).

There are reasons to assume that stem allomorphy can also contribute to inflectional uncertainty. If one verb has a gap in its paradigm, but the paradigm itself does not contain stem alternations, then it should be easier for a language user to predict an expected form for this gap than for a verb that has some stem alternation (of course it depends on how common/rare the stem alternation is in the first place). From 59 verbs we chose for the present study, 24 have a stem alternation realized in qualitative or quantitative consonant gradation. Quantitative consonant gradation (CG2) in Finnish is typically considered as a less complex (more predictable) allomorphy for language users than qualitative consonant gradation (CG1) (e.g., Karlsson 1983). CG2 is phonological in nature (read: automatic), meaning that speakers pronounce a long consonant (with a long voice onset time) in open syllables and a short version of the same consonant (with a short voice onset time) in closed syllables. There are only isolated words where these phonological correspondences do not apply consistently and these are treated as exceptions (see, Abondolo 1998; cf. to a similar set of correspondences that was completely morphologized in the Estonian language).

Smolek and Kapatsinski (2018, see also Stave et al. 2013) reported experiments where they created an artificial language in which words required certain consonant alternations when inflected. The authors claim that more complex alternations tend to be removed as the language is passed down because the likelihood increases over time that they will lose their productivity (cf. also Bybee 2008, who argued that as the magnitude of alternations (the number of articulatory changes) increases, their productivity decreases). Since changing the consonant quantity is of a lower magnitude than changing its quality, the magnitude of consonant gradation in Finnish seems to be entangled with the productivity of consonant gradation. Therefore, the magnitude of consonant changes and their productivity contribute orthogonally to uncertainty in case of defective paradigms. However, Smolek and Kapatsinski (2018) found a dissociation between production and judgment tasks in relation to the complexity of learned consonant changes. In their experiments, large changes that were judged as correct by listeners nevertheless tended to be avoided by speakers. For the current study this predicts that defective word forms that require qualitative consonant changes would be avoided by many participants in a production task and would be judged as plausible by many participants in a comprehension (judgment) task.

Do (2018) studied how four- to seven-year-old Korean speaking children unlearn the so-called output-output correspondence/faithfulness constraint in their initial grammar (the author uses Optimality theory framework), or, in other words,

how they learn stem allomorphy. The output-output correspondence constraint – if applied solely – would lead to paradigm uniformity (see, e.g., Kiparsky 1978). However, because of morphophonological changes in diachrony, children must incorporate knowledge of phonological alternations. According to Do's (2018) experiment results, children behave this way in experimental conditions that force them to favor alternations; however, they still show preference for paradigm uniformity whenever it is possible to avoid alternations. Because in Finnish, paradigmatic gaps presumably allow the avoidance of an expected word form by, e.g., substituting a synonymous form, we predict that those gaps that require stem alternations (especially qualitative consonant gradation) would be avoided by our participants. In a way, this behavior would also lead to paradigm uniformity, although by means of not-filling paradigmatic gaps and thus reducing the size of a paradigm. Therefore, we can assume that some paradigmatic gaps could potentially contribute to (or rather be a consequence of) paradigm uniformity.

The variable Consonant gradation was quantified as 0 (no gradation), 1 (quantitative and thus less complex), or 2 (qualitative and thus more complex).

2.3.2 Corpus frequencies

Before collecting the data described above (contemporaneity / subjective frequency / acceptability ratings for verbs, and information as to whether the verb undergoes qualitative or quantitative consonant gradation), we collected the following lexical variables: The frequencies of the lemma, past tense, and past participle forms of each verb were extracted from the Language Bank (of Finland) corpus (Aller Media Ltd. 2014), which includes 84.3 million word tokens from internet threads in a Reddit-like community. Neighborhood density for the verbs was calculated from the Basic Dictionary of Finnish (1990/1994) by counting the number of words with the same length but differing in their initial letter. Since Finnish orthography–phonology mapping is isomorphic, in the present study orthographic neighbors are equivalent to phonological neighbors. Bigram frequency, initial trigram frequency, and final trigram frequency (i.e., the average number of times that all combinations of two or three subsequent letters occur in the corpus) for verbs were obtained from the Turun Sanomat Corpus (22.7 million word tokens) using a computerized search program (Laine and Virtanen 1999). For contemporaneity we did not collect corpus frequencies, because the corpora of old(er) Finnish available are not of a size that would allow us to show reliably that the use of a lexeme was more frequent or central in the past.

2.4 Data analysis

All the analyses were produced in the statistical software R (R Core Team 2020). We used two statistical methods, one in which there is no dependent variable, and which allowed us to study relations between all the variables without giving special status to any one of them (Network analysis, NA, see, e.g., Borgatti et al. 2009), and another – linear mixed-effects models – that has rapidly become a gold standard for analyzing behavioral data (see, e.g., Baayen et al. 2008), and which is intended to explain the behavior of one variable (dependent) based on a group of other (independent) variables. In what follows, we describe these two methods in more detail.

We used two types of NA (both originate from graph theory, Newman 2010), first, to visualize our participants' inflectional choices (see Section 3.1 and Figure 2,

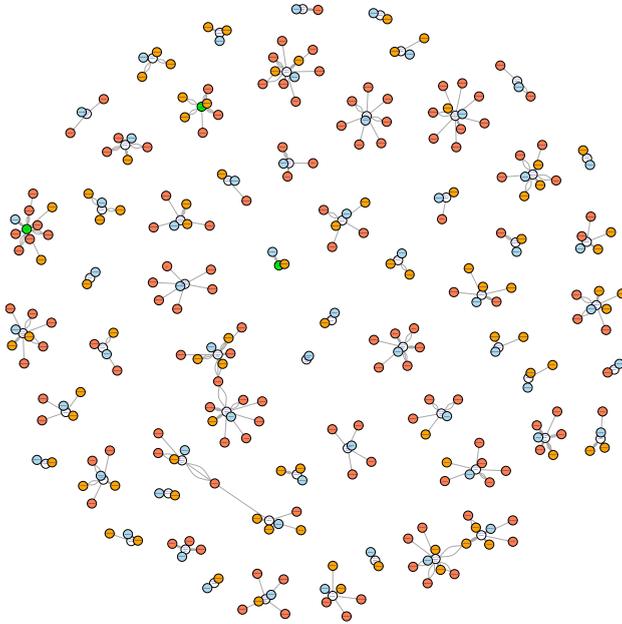


Figure 2: Network of inflectional choices. Each cluster of nodes (depicted by circles) represents one verb and consists of one ‘source’ node (depicted in white) and one or more ‘response’ nodes (depicted in light blue, orange, or coral, cf. Figure 1) connected by edges (lines with arrowheads). White nodes depict the past tense forms that participants saw; colored nodes represent participants’ inflectional choices. White nodes also contain averaged contemporaneusness/frequency/acceptability ratings given by native speakers who did not take part in the verb production task. In addition, white nodes contain information as to whether the verb stem undergoes qualitative (CG1) or quantitative (CG2) consonant gradation. This information is also presented in the Supplementary materials (pp. 14–23). Online readers can zoom in on the figure in order to see the nodes more clearly.

for which we used the package *igraph*, Csardi and Nepusz 2006), and second, more importantly, to analyze statistically how the variables from all the data sources (verb production and evaluation tasks as well as corpus frequencies and grammatical characteristics of verbs such as consonant alternations) are related to each other (in other words, correlated) in the network and also to visualize these relations (see Section 3.2 and Figure 3). Networks consist of nodes (typically depicted by circles) and edges (lines) that connect nodes. Nodes could represent inflectional variants the participants produced (Figure 2), or they could represent variables, e.g., corpus frequencies of these inflectional variants (Figure 3). The network shown in Figure 2 is a so-called unweighted network, because there is no information about the intensity of a relation between connected nodes. However, the network showed in Figure 3 visualizes the intensity of relations between, e.g., Lemma corpus frequency and Past tense corpus frequency by the thickness and the length of the edge between the two nodes. Positive associations between the variables are depicted by blue edges, and negative ones by red edges, and they are calculated using Spearman's rank correlation coefficient. Thus, the nodes in

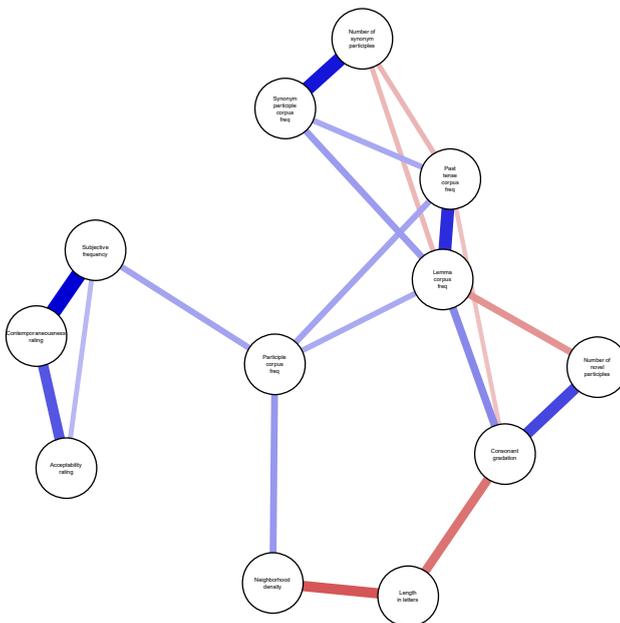


Figure 3: Estimated network of predictors. Blue edges (lines) indicate positive relations (increase in A related to increase in B) while red edges indicate negative relations (increase in A related to decrease in B). The size and the color intensity of edges show the intensity of the relationship.

Figure 3 represent observed variables, connected by edges estimated from data and thus representing statistical relationships. The network analysis for Figure 3 was calculated in the package *bootnet* (Epskamp et al. 2018) using the function *estimateNetwork*. We used the least absolute shrinkage and selection operator (LASSO, Tibshirani 1996) to obtain a conservative (sparse) network model with only a relatively small number of edges to explain the covariation structure in the data (Jankova and Van de Geer 2018). A tuning parameter was selected by minimizing the extended Bayesian information criterion (EBIC; Chen and Chen 2008).

Another difference between the networks in Figures 2 and 3 is that in the latter the relationship between the variables is symmetrical, hence the edges are undirected, whereas in the former the relationship is not symmetrical, and the edges are directed (depicted by arrowheads which represent temporal dependencies: one word form was shown to a participant on a computer screen, and another word form of the same lexeme was produced by the participant after he/she saw the presented word form).

Figure 4 was produced in the package *qgraph* (Epskamp et al. 2012) using the function *centralityPlot*. It reports the so-called centrality indices for NA, which quantify the relative importance of a node in relation to other nodes in the network depicted in Figure 3 (see Costantini and Perugini 2020). We report the *strength* of the interactions that a node has with its neighbor nodes (again, each node represents a variable, not a word form or a participant), *betweenness*, i.e., the number of times a node lies on the shortest path between two other nodes, and *closeness*, or how well a node is (in)directly connected to other nodes. In each of these definitions of centrality, a node can be somewhere on the continuum from central to peripheral and the three indices are sorted according to *closeness*.

Contrary to classical latent variable models, which mainly focus on what is common among different lexical predictors, NA focuses on what is specific to lexical predictors in the context provided by other predictors in the network. Recurring information (or redundancy) is a fundamental property of language (Milin et al. 2016) and one of advantages of NA is that it does not require the removal of redundancy (collinearity) from predictors prior to modeling.

In the three linear mixed-effects models (Bates et al. 2015) we analyze/explain three dependent variables by the number of independent variables (predictors, see Section 3.3). The dependent variables stem from the data collected via the verb production task. In the first model (see Table 1), we used expected versus novel forms the participants produced as the response (dependent) variable. In the second model (see Table 2) we used expected versus synonymous forms the participants produced as the response variable. Both models were logistic regression models, or, in other words, generalized mixed-effects models. For the third model (see Table 4), we quantified this variation into one dependent variable: we

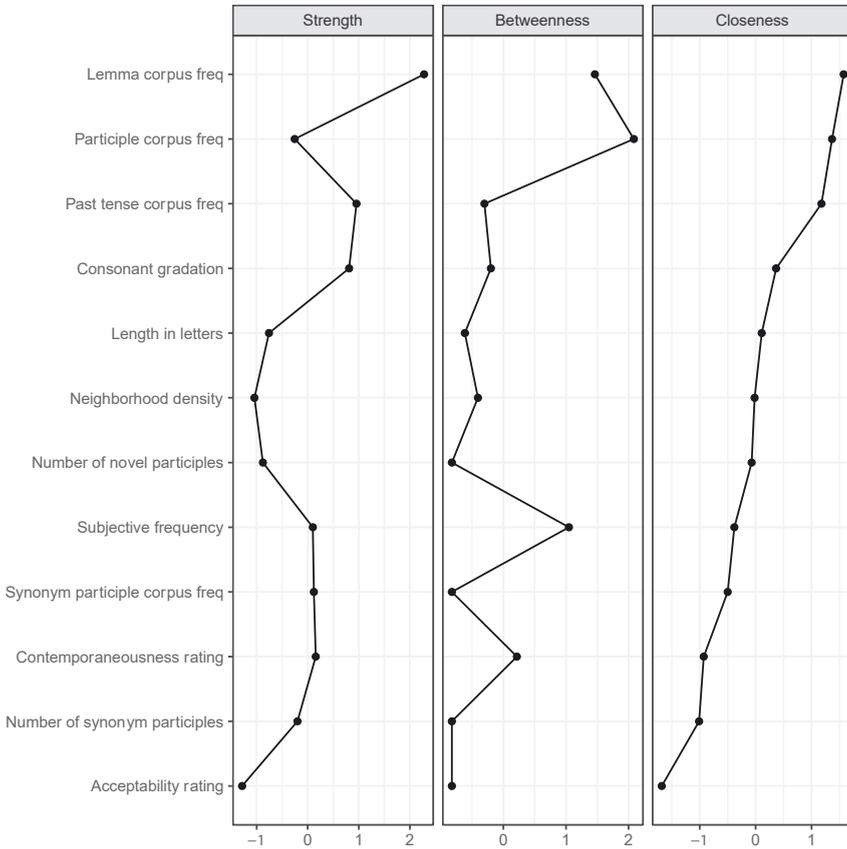


Figure 4: Centrality indices: **STRENGTH** (how well a node is directly connected to other nodes); **BETWEENNESS** (shows how important a node is on the average path between other nodes); **CLOSENESS** (how well a node is directly or indirectly connected to other nodes).

compared each answer of each participant to an “ideal” answer, which we called “expected” (see Figure 1), using the so-called Optimal String Alignment (OSA, van der Loo 2014) method (for a similar approach, see Nikolaev et al. 2020). The OSA is a method of comparing two words similar to the Levenshtein distance method (Levenshtein 1966). The Levenshtein distance is computed simply by counting the number of insertions, deletions, and substitutions necessary to turn one word into another. However, the difference between this method and the OSA is that the latter allows for transpositions of adjacent characters. The OSA is sometimes referred to as the restricted Damerau-Levenshtein distance (Wagner and Lowrance 1975; van der Loo 2014). To put it simply, we calculated how far away a real answer

is from an “ideal” answer: how many Levenshtein distance steps are required to turn a real answer to an expected answer. For example, if many steps are required (e.g., 4), this means that the produced and expected words are barely related to each other (although, cf. as a counter-example the two word forms *sneaked* and *snuck*, for which the OSA distance is 4).

To overcome another requirement of the mixed-effects models, namely independence of predictors (explanatory variables; that is why they are sometimes called independent variables), we collapsed all of them into principal components (for a similar approach, see, e.g., Nikolaev et al. 2019). Principal component analysis (PCA) is one of the methods of dimensionality reduction. In the Supplementary materials (pp. 24–26), we report PCs and the loadings sorted for each PC (the rotation matrices for these PCs) and we interpret PCs for those variables that have strongest negative or positive loadings on them.

We added participants’ gender, age, and years of education as participant-level explanatory variables. Together with principal components they constituted fixed effects (or in other words, population-level effects). We also added significant fixed effects as by-participant random slopes into the model (see Barr et al. 2013). Participants and words (experiment items) constituted random effects (or, group-level effects). In each model we added predictors one at a time and compared the model with a new predictor with the previous (simpler) model using the function *anova*. If the model with the additional predictor was significantly better (based on the p-value of the *anova* comparison as well as the lower AIC index of the new model, meaning a better fit), we kept the predictor.

3 Results

3.1 Verb production

As we expected, most of the 59 verbs triggered more than one inflectional choice. Figure 2 shows the results presented as a network to visualize inflectional choices. This network consists of nodes (depicted by circles) and edges (lines with arrowheads) that connect nodes. Each cluster of nodes represents one verb and consists of one ‘source’ node (depicted in white) and one or more response node (depicted in light blue, orange, or coral, cf. Figure 1). The more people chose one particular inflectional variant, the more edges this variant has and the closer it is to the source node. Arrowhead edges thus present dependencies: e.g., the participants read the form *rohkeni* ‘dared’ (the source node), from which they produced, e.g., the participle *rohjennut* (the response node, in this case depicted in light blue). The distance between connected nodes is meaningful, with longer distances representing

more sporadic connections while the distance between unconnected node clusters is not meaningful, i.e., node clusters next to each other are no more closely related than those at opposite ends of the diagram. Three verbs that are defective according to the switch model of paradigmatic defectivity are depicted with a green source node. However, as one can see, they do not seem exceptional compared to many other verbs that triggered a similar level of uncertainty in participants.

3.2 Network analysis

We quantified inflectional uncertainty as measured by the number of different synonymous or novel past participles produced for each verb and included these variables with other predictors (see Section 2.3) in the following network model (see Figure 3). We also retrieved from the corpus (84.3 million word tokens, Aller Media Ltd. 2014) the frequencies of those synonymous past participle forms that participants had used in their inflectional choices.

Unlike the previous network (Figure 2), which consisted of 59 separate clusters (participants' inflectional choices for 59 verbs), the network depicted in Figure 3 consists of nodes representing observed/collected variables for these verbs (e.g., lemma frequency, word length, neighborhood density etc.), connected by edges estimated from the data and thus representing statistical relationships between these variables (Epskamp et al. 2018).

Network analysis focuses on what is specific to observed variables in the context provided by other variables in the network. Figure 3 shows that similar nodes cluster together in terms of their effect on overall uncertainty. On the left side, the three ratings-task nodes are (as we expected) strongly connected to each other and weakly connected to the network of other features through the node *Participle corpus freq.* Since the participants were asked to give their subjective ratings of the participle forms, it makes sense that their subjective frequency ratings are connected to the network through the corpus frequencies of these participle forms. The fact that contemporaneousness and acceptability ratings do not have their own edges connecting them to the network shows that these two ratings measure somewhat different (albeit related) concepts than the subjective frequency rating does. In addition, within this triangle of subjective measures, acceptability of the form that is supposed to fill a paradigmatic gap is less related to this form's perceived frequency than its perceived archaicity is. Therefore, acceptability in a word comprehension task is only weakly related to acceptability in a word production task: even if one can accept more easily some suggested form to fill a gap in a paradigm, one would not necessarily come up with this form himself/herself when asked to fill a gap.

In the center/upper right, corpus frequencies have some strong connections to each other as well as to the node *Number of synonymous participles*. By synonymous participles, we mean participles formed from those verbs that a) are different from our target verbs and belong to a different inflectional type from the target verbs, b) are listed in the Basic Dictionary of Finnish, and c) share their meaning with the target verbs. Before estimating the network (Figure 3), we retrieved corpus frequencies for each of the synonymous past participle forms produced by participants (for 48 out of 59 verbs, participants produced at least one and at most three different synonymous participles) and included them in the network (node *Synonymous participle corpus freq.*). As can be seen from Figure 3, this node is strongly connected to inflectional uncertainty as measured by the number of synonymous participles (node *Number of synonymous participles*), meaning that if there exist synonymous verbs from other, non-defective inflectional types, people tend to substitute these synonymous participles for defective-type participles and the number of such participles attested positively correlates with the frequency of the synonymous verbs.

Both measures of inflectional uncertainty, *Number of synonymous participles* and *Number of novel participles*, are connected to lemma frequencies calculated from the corpus (Aller Media ltd. 2014). The higher the lemma frequency of the verb, the lower the number of novel or synonymous participles produced. Lemma frequency is also connected to other corpus frequency measures: frequencies of participle forms (expected forms), frequencies of past tense forms (those that participants read in sentence #1, see Section 2.1.1), and frequencies of synonymous participle forms (see previous paragraph). In other words, a participle or a past tense form of a verb with high lemma frequency tends to be of high frequency too, and it also tends to have synonyms with high frequency (and vice versa: replace the word *high* with the word *low*). The first part of the previous sentence should not surprise the reader (if a verb is frequent, then on the whole, its individual forms tend to be frequent too), but the second part (frequency of synonyms) is not self-evident and it needs an explanation (see Discussion).

In the lower right, structural features all have strong connections with each other. Inflectional uncertainty as measured by the number of novel participles is strongly connected to the stem allomorphy of verbs (node *Consonant gradation*). Verbs with consonant gradation in their stems are associated with a higher number of novel past participles produced by participants than verbs without consonant gradation. Within verbs with consonant gradation, more complex stem alternation pattern of the verb (less predictable, because it is a relic of historical changes) tends to positively correlate with a higher number of novel past participles produced by participants. *Length in letters* is negatively connected to *Consonant gradation* and to *Neighborhood density*. The longer the word, the lower the chances

of having phonological neighbors. Gradation is a historical process applied to proto-languages that are conventionally reconstructed as having a predominantly disyllabic word stock. There is also presumably a negative correlation between word length (possibly above a disyllabic floor) and type frequency, so there are fewer potential neighbors for longer words.⁵ Words with consonant gradation in the inflectional type we chose for the present study were on average shorter than words without consonant gradation (5.96 vs. 6.71 letters, $t = 3.433$, $p = 0.0013$). The strong correlation between length and phonological density measures such as neighborhood density (calculated by counting the number of words with the same length but differing only in the initial letter; used in the current study) or Hamming distance of one (calculated by counting the number of words with the same length but differing in any single letter; not used in the current study) can be reduced by using alternative measures (see, e.g., Kapatsinski 2006). Because we do not have any specific hypotheses regarding phonological or orthographic neighborhood density and its influence on paradigmatic gaps, we leave these potentially important measures for future studies.

Bigram and trigram frequencies do not correlate significantly with any other variable in the network, so they are not present in the estimated network.

In Figure 4 we report the so-called centrality indices, which typically quantify the relative importance of the node in relation to other nodes in the network (Costantini and Perugini 2020). We report *STRENGTH* (the strength of the interactions that a node has with its neighbor nodes), *BETWEENNESS* (the number of times a node lies on the shortest path between two other nodes), and *CLOSENESS* (how well a node is directly or indirectly connected to other nodes).

Figure 4 shows that the *Acceptability rating* provided by native language users for past participle forms is on the extreme left of the continuum from peripheral to central in each of these definitions of centrality, meaning that this variable is peripheral in the estimated network. On the extreme right of the continuum in each of these definitions of centrality is *Lemma corpus frequency*, emphasizing its central status in the estimated network. In other words, nodes with more connections are more central to the structure of the network and tend to have a greater capacity to influence other nodes (Yan and Ding 2009). Therefore, of all our factors, Lemma corpus frequency appears to be the most collinear with other predictors and closer to the number of synonyms or novel forms produced (negative correlation) than acceptability ratings are to these indices of uncertainty.

⁵ We thank an anonymous reviewer for suggesting this explanation of the correlation between length, gradation, and neighborhood density.

3.3 Linear mixed-effects models

In this section, we present the results of three linear mixed-effects models we ran in order to explain the individual inflectional choices (59) made by each participant (50). Thus, we have a table with 2,949 responses⁶ (59×50).

The response variable in the first model (see Table 1) was categorical: EXPECTED (coded as 1) versus NOVEL FORM (coded as 0). The response variable in the second model (see Table 2) was also categorical: EXPECTED (coded as 1) versus SYNONYMOUS FORM (coded as 0). The set of predictor variables are the result of the PCA analysis described in the Supplementary materials (pp. 24–26).

In the model explaining expected and novel word forms (Table 1), two predictors turned out to be significant. PC1 has a positive estimate, showing that subjective frequency, contemporaneousness, and acceptability ratings of expected participle forms significantly explain the number of expected participle forms our participants produced. The higher the subjective ratings of a form, the more likely participants were to produce this form. The lower the ratings, the more likely the participants were to fail in their attempt to produce an expected form (as a result they produced novel forms). PC3 with a strong positive loading of stem alternation, has a positive estimate. This demonstrates that the participants were more likely to produce a novel participle form for words with stem alternation (consonant gradation) and an expected word form for words without stem alternation.

Table 1: Estimated coefficients, standard errors, *z*- and *p*-values for the generalized mixed-model fitted to the categorical variable EXPECTED (coded as 1) versus NOVEL FORM (coded as 0) for 59 verbs in the verb production experiment.

| Fixed effects | Estimate | Std.Error | <i>z</i> -value | <i>p</i> -value |
|----------------|-------------|-----------|-----------------|-----------------|
| (Intercept) | 2.369 | 0.289 | 8.208 | <0.001 |
| PC1 | 0.633 | 0.151 | 4.204 | <0.001 |
| PC3 | 0.619 | 0.204 | 3.035 | 0.002 |
| Random effects | | | | |
| Groups | Name | Variance | Std.Dev. | Corr |
| Item | (Intercept) | 2.384 | 1.544 | |
| Subject | (Intercept) | 0.665 | 0.815 | |
| | PC1 | 0.029 | 0.169 | 0.35 |
| | PC3 | 0.025 | 0.159 | 0.21 |

Number of obs. 1,905; Items, 59; Subjects, 50.

⁶ One response was missing and was coded as NA.

Table 2: Estimated coefficients, standard errors, *z*- and *p*-values for the generalized mixed-model fitted to the categorical variable EXPECTED (coded as 1) versus SYNONYMOUS FORM (coded as 0) for 59 verbs in the verb production experiment.

| Fixed effects | Estimate | Std.Error | <i>z</i> -value | <i>p</i> -value |
|---------------|----------|-----------|-----------------|-----------------|
| (Intercept) | 1.538 | 0.395 | 3.894 | <0.001 |
| PC1 | 1.204 | 0.28 | 4.304 | <0.001 |
| PC3 | -1.037 | 0.241 | -4.297 | <0.001 |
| PC5 | -1.914 | 0.567 | -3.377 | <0.001 |

| Random effects | | | | | | |
|----------------|-------------|----------|----------|-------|-------|------|
| Groups | Name | Variance | Std.Dev. | Corr | | |
| Item | (Intercept) | 3.996 | 1.999 | | | |
| Subject | (Intercept) | 2.638 | 1.624 | | | |
| | PC1 | 0.043 | 0.207 | 0.66 | | |
| | PC3 | 0.025 | 0.158 | -0.31 | -0.92 | |
| | PC5 | 0.05 | 0.223 | -0.08 | -0.80 | 0.97 |

Number of obs. 2,537; Items, 59; Subjects, 50.

The sample we used for the models presented in Tables 1–3 is reasonably large (50 participants), however, to complement these frequentist models, we also replicated them by fitting Bayesian generalized linear multivariate multilevel models,⁷ as the Bayesian credible intervals are sensitive to the amount of data (power).

In the model explaining expected and synonymous word forms (Table 2), three predictors turned out to be significant. As in the previous model, PC1 has a positive estimate, showing that subjective frequency, contemporaneity, and acceptability ratings of expected participle forms significantly explain the number of expected participle forms our participants produced. The higher the subjective ratings of a form, the more likely participants were to produce this form. However, contrary to the previous model, PC3 with a strong positive loading of stem alternation has a negative estimate. This demonstrates that the participants were more likely to produce a synonymous participle form for words without stem alternations in comparison to words with the stem alternation. We return to this surprising finding in the Discussion. PC5 was another significant predictor with a negative estimate in the model. PC5 reflects corpus frequencies of expected participle forms (with a loading of -0.70), and corpus frequencies of synonymous participles the participants produced (with a loading of 0.41). This principal component thus

⁷ The number of Markov chains = 4 and the number of total iterations per chain including warmup = 4,000. We used the R package *brms* (Bürkner 2017, 2018, 2021). The script and the data to re-run the models is available here: <https://doi.org/10.17605/OSF.IO/T278K>.

predicts that the higher the corpus frequency of the expected participle form, the more likely participants are to produce this expected form (and vice versa: the lower the frequency, the more likely they are to produce a synonymous word form). In addition, PC5 predicts that the higher the frequency of any of synonymous routes by which a participant can substitute an expected form, the more likely the participant is to do that.

The response variable in the third model (see Table 4) was the OSA distance between expected and produced participle forms described in Section 2.4. Since the OSA distance for expected forms is zero, we included in this model only novel and synonymous responses. The OSA range of the variants produced was 0–7. Actual frequencies of each type with total percentage are presented in Table 3.

Table 3 shows that the ratio between expected and other participle forms (synonymous or novel) the participant produced is approximately 1:1. The vast majority of synonymous forms (857/1,045, 82%) had an optimal string alignment distance of two letters/phonemes from the expected form (this is possible because the expected and synonymous verbs are typically derived from the same adjective or noun, with the only difference being that they have different derivational suffixes).

Novel forms were more evenly distributed across the scale, but even so, the majority (62.1%) were one or two places distant from the expected form. Many novel past participles produced by the participants are similar to synonymous participles, since they often use the same derivational suffix; however, we do not regard them as synonymous, because they are based on derived verbs that are not in conventional use and not listed in dictionaries.⁸ The OSA distance of one could

Table 3: Total numbers of expected, novel, and synonymous past participle forms produced by participants according to the Optimal String Alignment (OSA) distance between the expected and produced forms (with relative percentages among totals).

| OSA | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------------------|------------------|---------------|------------------|--------------|---------------|--------------|-------------|-----------|
| Expected forms | 1,492 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Novel forms | 0 | 103 | 162 | 49 | 81 | 13 | 4 | 0 |
| Synonymous forms | 0 | 7 | 857 | 21 | 152 | 3 | 4 | 1 |
| Total | 1,492 (50.6%) | 110 (3.7%) | 1,019 (34.6%) | 70 (2.4%) | 233 (7.9%) | 16 (0.5%) | 8 (0.3%) | 1 (0%) |

⁸ For example, the OSA distance between the expected past participle form *toennut* of the verb *toeta* ‘calm down’ and the novel past participle form *toentunut* (produced by one participant) of the verb *toentua* is 2. However, the verb *toentua* is not mentioned in any dictionary nor found in the corpus we used, even though this novel verb has the same derivational suffix *-ntu/nty* as many synonymous verbs.

be attributed in some cases to failed attempts to choose an appropriate stem allomorph due to consonant gradation (cf. the novel form *rohkennut* for the expected *rohjennut*).

In the model explaining the OSA distance between expected and produced inflectional choices (Table 4), three predictors turned out to be significant.

First, the distance between synonymous forms and expected forms is greater than between novel forms and expected forms, which is a predictable result, since novel forms can be considered as non-conventional (or failed?) attempts to produce an expected form, while synonymous forms represent a conventional (read: existing) inflectional type which is different from an expected inflectional type and thus these synonymous forms have greater chances to be more distant (in OSA distance terms) from the expected forms. When respondents create their own (non-conventional/non-existent) inflectional variant, they cannot be too creative (cannot introduce too many changes), otherwise they will be sanctioned by other language users. When a respondent follows an existing inflectional route (albeit a different one as in the case of using synonymous verbs), he/she is producing a conventional (existing) form – no matter how distant it is from the expected form – and thus will not be sanctioned by other language users.

Second, PC1 has a negative estimate, showing that subjective frequency, contemporaneousness, and acceptability ratings significantly explain the OSA distance between expected and produced participle forms; therefore, higher

Table 4: Estimated coefficients, standard errors, *t*- and *p*-values for the mixed-model fitted to the OSA distances between expected and produced participles for 58 verbs in the verb production experiment.

| Fixed effects | | | | | | |
|----------------|-------------|----------|-----------|-----------------|-----------------|-------|
| | | Estimate | Std.Error | <i>t</i> -value | <i>p</i> -value | |
| (Intercept) | | 2.054 | 0.104 | 19.809 | <0.001 | |
| Synonym | | 0.315 | 0.086 | 3.649 | <0.001 | |
| PC1 | | -0.12 | 0.044 | -2.698 | 0.009 | |
| PC3 | | -0.223 | 0.047 | -4.72 | <0.001 | |
| Random effects | | | | | | |
| Groups | Name | Variance | Std.Dev. | Corr | | |
| Item | (Intercept) | 0.145 | 0.38 | | | |
| Subject | (Intercept) | 0.282 | 0.531 | | | |
| | Synonym | 0.162 | 0.403 | -0.97 | | |
| | PC1 | 0.013 | 0.115 | 0.02 | -0.19 | |
| | PC3 | 0.008 | 0.087 | -0.43 | 0.34 | -0.21 |
| Residual | | 0.473 | 0.688 | | | |

Number of obs. 1,457; Items, 58; Subjects, 50.

ratings (more familiar, more contemporary, more acceptable forms) predict less distance between expected and produced forms (and vice versa, lower ratings predict greater distances). In other words, expected participle forms with higher ratings seem to restrict possible inflectional choices so that the produced participle forms tend to be closer to expected ones, while participle forms with lower ratings open more possibilities to maneuver.

Third, PC3, with a strong loading of stem alternation, has a negative estimate. Since consonant gradation has a negative loading (-0.65) on PC3, this demonstrates that consonant gradation in the verb stems is positively associated with the OSA distance between expected and produced participles. Verbs with no consonant gradation tend to predict less or no distance, while verbs with consonant gradation tend to predict a greater distance. Similarly, verbs with less complex (quantitative) consonant gradation tend to predict less distance than verbs with more complex (qualitative) consonant gradation. However, PC3 is not as informative in this model as in the models presented in Tables 1 and 2. The alternative forms are likely not to have stem alternations, so if the target form has a stem alternation, they could just automatically become farther from it.

PC2, PC4, bigram and trigram frequencies, age, gender, and years of education were not significant predictors of participants' inflectional choices in our models.

4 Discussion

We argue that paradigmatic defectivity can be described as a dynamic rather than a static system, where gaps represent a continuum of possible inflectional choices rather than simply a lack of an inflectional variant. However, this continuum has a cognitive tradeoff between offering a solution to a problem and imposing a cost related to this solution. If we describe gaps as portals to other paradigms that help language users to escape an unproductive paradigm, there is still some pressure to form an expected word form (as was demonstrated in our study), with or without success. And while doing that, language users have to deal with a heavier cognitive load due to processes such as word retrieval, word composition, activation of several competing forms, and inhibition of less suitable variants.

One common strategy of word inflection in situations when a word does not have a full inflectional paradigm (when some word forms are missing) is to produce an expected word form, which results in either success (an expected form was produced to fill a gap) or failure (a novel word form was produced to fill a gap). If respondents have never heard/read/produced this particular word form before, they could deduce the form by analogy with a similar paradigm using some other member of the paradigm as a starting point. When participants of the current study tried to deduce a past participle form by analogy to words with a similar paradigm using as a starting point the third singular or plural past tense form that was

presented to them, this strategy failed in some participants for some words, and as a result they produced novel forms. Mixed-effects models (see Section 3.3) showed that the lower the subjective frequency, contemporaneousness, and acceptability ratings of expected participle forms, the more likely the participants were to produce novel participle forms. We posited in the Introduction that success or failure would depend on the complexity of the paradigm (a more complex paradigm will cause more failure). Indeed, this is what we found in the model comparing expected and novel forms. Likewise, our network analysis (see Section 3.2) showed that the number of different novel forms produced is related to consonant gradation of verbs. Failure of the strategy for producing an expected word form is therefore positively associated with an increase in structural complexity expressed in stem alternations. Our findings regarding an effect of consonant gradation are in line with those reported in Hedlund et al. (2021), in which Finnish native speakers inflected real nouns and pseudo-nouns with or without consonant gradation. Hedlund et al. (2021) found that a) pseudo-nouns yielded higher error rates than real nouns, and b) within pseudo-nouns those that required consonant gradation resulted in more errors. However, Hedlund et al. (2021) do not differentiate between qualitative and quantitative gradation.

Another common strategy for word inflection in situations when a word does not have a full inflectional paradigm is avoidance: language users simply replace a missing word form with another word form (from a full paradigm) when the word conveys a similar meaning. The lower the subjective frequency, contemporaneousness, and acceptability ratings of expected participle forms, the more likely the participants were to produce word forms from synonymous paradigms to fill a paradigmatic gap. Production of expected versus synonymous forms depended on the complexity of the paradigm; however, counter-intuitively, verbs with more complex paradigms resulted in fewer synonymous word forms. The complexity of the paradigm is thus related to inflectional strategies a speaker chooses; however, it depends on whether the form exists (is in conventional use as in the case of synonyms) or not (is not in conventional use as in the case of novel forms). A structural (form-based) and semantic (meaning-based) similarity might have caused competition between the two synonymous words. In the long run, at a certain point, the word with a less complex (more predictable) inflectional paradigm started winning this competition. Hence, verbs with more complex paradigms (which include unproductive, frozen stem alternations) seem to offer more paths to words with the same meaning but a different form (and hence paradigm). Regarding the losing word, this decrease in use was probably not equally distributed among all members of its inflectional paradigm. Those members that fell out of use first started to be perceived as gaps in the paradigm and the paradigm can thus come to be described as defective. Since this is a process with no abrupt start or end, the present study is only a snapshot on a timeline of one inflectional verb class in Finnish.

We also found that the synonymous word forms tend to be more distantly related form-wise to expected forms than the novel ones. The reason for this is that novel word forms are failed attempts to produce an expected word form to fill a gap. Our results show that these failed attempts tend to be close (or at least closer) to the target forms than when a speaker chooses a synonym and thus a different paradigm. Synonymous forms are in conventional use and thus cannot be “sanctioned” by interlocutors, no matter how distant they are form-wise from expected forms, while novel forms that our participants produced to fill paradigmatic gaps seem to have fewer degrees of freedom: the farther away the novel form is from an expected one, the higher the potential “sanctions”.

In the Introduction, based on findings by Harmon and Kapatsinski (2017) about frequent forms’ extension (a frequent form is likely to be chosen over a less frequent synonym), we predicted that the number of synonymous forms produced to fill paradigmatic gaps would negatively correlate with the lemma frequency of the words containing paradigmatic gaps and would positively correlate with frequency of these synonymous words. The network analysis (see Section 3.2) confirmed this prediction: frequent forms are extended to novel uses (via synonymy) because they are more accessible than their competitors (Harmon and Kapatsinski 2017; Zipf 1949). It also showed that the number of different synonyms our participants produced tends to positively correlate with the frequency of their lemmas. Since this is not a self-evident fact, one possible line of reasoning is the following: we have a limited set of 59 verbs. Within these, the general lexical field expressed by the less common verbs is in fact less needed, less used, so their synonyms will also be less frequent. If we take as an example two Russian verbs with defective paradigms, the verb *победить* ‘conquer/win’ (16,480 tokens in the Russian National Corpus (RNC), www.ruscorpora.ru) and the verb *пылесосить* ‘vacuum’ (81 tokens in the RNC), then *победить* has broader application and will have more frequent synonyms than *пылесосить*, which is lexically quite specific. This might be true of our particular sample without being necessarily true of the language at large; however, the effects of lexeme frequency on synonym substitutions (Harley and MacAndrew 2001; Kittredge et al. 2008) and replacement repairs (Kapatsinski 2010), which can involve semantically similar words, suggest support for this interpretation, especially as it does not seem to be the case for phonological speech errors.⁹

Contemporaneousness, subjective frequency, and acceptability ratings of participle forms strongly correlate with each other; however, they seem to correlate less with the two measures of uncertainty (the number of novel participles and the

⁹ The authors would like to thank Vsevolod Kapatsinski (personal communication) for this observation.

number of synonymous participles the participants produced) than some other variables we used in the network (such as the lemma frequency, i.e., the summed frequency of all the paradigm members for a lexeme, and stem alternations that are present or absent in the paradigm of the lexeme). One possible explanation for this relatively weak connection between uncertainty observed in a production task and subjective measurements collected from a comprehension task might be in the nature of these two tasks. When one has to produce a form to fill a gap, one's uncertainty is of a different magnitude than when judging how frequent, archaic, or acceptable certain form is. Another possible explanation of a weak connection between subjective ratings and uncertainty is that the network analysis is done using aggregated data. The data has 59 rows, as many as there are test items in the production or comprehension tasks. Therefore, the two measures of uncertainty (how many different synonyms and how many different novel forms were produced for each word) in this type of analysis are both at the collective level, not individual. However, when we shift from the collective level to analyzing uncertainty at the individual level (explaining everyone's inflectional choice using mixed-effects models), then the perceived archaicity, lower frequency and lower acceptability of word forms are strongly associated with the relative unpredictability of their paradigms (seen in inflectional choices the participants made).

For 58 out of 59 verbs, an expected past participle was produced at least by one participant (nobody produced an expected participle from the past tense form *erkani* 'separated'). For only one out of 59 verbs did all participants produced only the expected past participle (*auennut* 'opened', see Figure 2). According to the switch model of paradigmatic defectivity, only one verb should be regarded as defective (the verb *erata* 'separate'), all other verbs being not defective. According to the usage-based (dimmer) model of paradigmatic defectivity, one verb (the verb *aueta* 'open') in this inflectional type is likely not to be defective and all other verbs are likely to be regarded as defective. We used the phrases *likely to be defective* or *likely not to be defective* in the previous sentence since the usage-based (dimmer) model of defectivity is not deterministic. Even though the verb *aueta* 'open' was not defective in our study, there is always some small possibility that it could be defective to some people if we used a larger sample of participants. The opposite is also true: there could be one person in a sample who produces a different-from-expected form for a verb while all other participants produce an expected form only. How much weight should we put on one person's decision? Are we allowed to conclude that the verb has a defective paradigm based on this kind of data (e.g., one unexpected response out of the 50 expected)? What if this was just a typo, a slip of the tongue, a slip of the mind? The nondeterministic nature of the usage-based (dimmer) model of defectivity is in

line here with Bayesian thinking: we could simply let the data update our priors (expectations), e.g., that the paradigm of the verb *aueta* ‘open’ is not defective, without ever assigning our priors to either zero or one.

According to Goldberg (2019), language users consider creative (novel) uses “wrong” when there exists a conventional (expected) alternative way to express the same meaning, because they tend to view language normatively: naïve speakers consider there to be “right” ways to use language (cf. one of our participants asked us to send her the ‘correct’ past participles after she finished the verb production task, so that she could compare her answers to the “correct” ones). Paradigmatic gaps are islands of freedom, where there seems not to be a conventional (expected) word form preempting creative use. And yet, despite these suddenly increased degrees of freedom, many people in the present study managed to produce conventional (expected) forms (one participant produced only 12 out of 59 expected forms while two participants produced as many as 56 out of 59 expected forms (mean 29.8, SD 11.3), see Figure 5).

In Figure 5, the switch model of defectivity would predict a flat contour so that each bar (corresponding to each participant) would have a height of, e.g., 56, suggesting that there are only three verbs in this inflectional type that are defective all other verbs being non-defective. However, Figure 5 instead reflects the absence of consensus in a speech community, thus supporting the dimmer model of defectivity. Age, gender, and years of education are not significant factors that would explain this individual variation (or we need to widen the range for age and years of education). Since (at least partially) creative versus conventional language behavior seems to be sociolinguistic in nature (Goldberg 2019: 61), further research is needed to understand why some people tend to be more creative while some other people tend to be more conventional in their language use (for more detailed discussion of the social basis of language, see Divjak et al. 2016, and Langacker 2016).

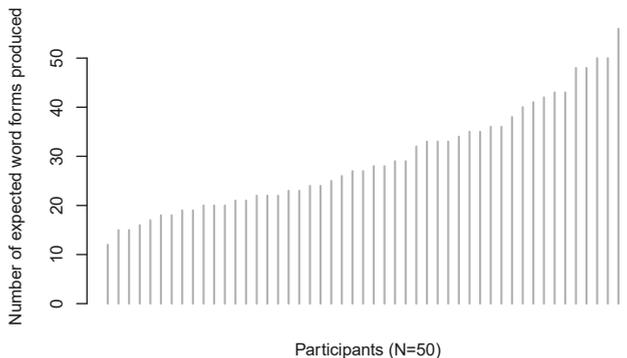


Figure 5: The number of conventional (expected) word forms (y-axes) the participants (x-axes) produced for 59 Finnish verbs belonging to the same inflectional type.

Summa summarum, the dimmer (usage-based) model of paradigmatic defectivity that we discussed in the introduction fits better with results of the current study than the classical switch version of paradigmatic defectivity typically met in dictionaries and descriptive grammars.

5 Conclusion

The purpose of language is not to produce different forms, but rather to convey different meanings and functions, which may entail employing a variety of forms. However, in inflectional languages, some words, along with their inflectional paradigms, become archaic, while other words join more productive inflectional classes, and occasionally the reverse can happen as well. No specific change is inevitable, as all changes are usage driven. This protracted competition can lead to paradigmatic defectivity, where it is no longer automatic to determine what a suitable form is for a given slot. As a result, the relation between complexity of the form and complexity of the meaning is no longer linear. E.g., the meaning of the Russian equivalent of the phrase *I'll win* is no more or less complex than the meaning of the Russian equivalent of the phrase *you'll win*, although the form (a slot in the inflectional paradigm of the verb *win*) for the previous does not seem to exist (?я победю/?я побежду) while the form for the latter does exist (ты победишь 'you'll win').

The variety of responses in our study suggests that the manner in which a certain word is inflected is driven by the producer's experience with the word's paradigm and with related paradigms, and that there are a variety of strategies that language users employ to this end. One strategy of dealing with inflectional uncertainty related to words with defective paradigms was seen above: to choose another word from a rival inflectional type that conveys similar meaning (a process that leads to the weakening of one inflectional type and the strengthening of another). Another common strategy of dealing with paradigmatic defectivity was to remain within the chosen defective paradigm and try to predict (using analogy) a word form that one has never met before. The success or failure of this strategy depends, we propose, on the predictability of the paradigm, which in turn is related to a word's lemma frequency and is partially related to the morphophonological complexity of its paradigm. More complex paradigms (those with more morphophonological changes throughout the paradigm) tend to be less predictable, and this may prove to be the reason why the paradigm loses in competition with other paradigms and becomes defective or obsolete.

Data availability statement

The datasets generated and analyzed during the current study are available in the Open Science Framework repository, at the following link: <https://doi.org/10.17605/OSF.IO/T278K>.

Acknowledgments: We thank all our participants. We also express our gratitude to Vsevolod Kapatsinski and an anonymous reviewer for their constructive comments and suggestions.

Research funding: The authors gratefully acknowledge the support of the UK Arts and Humanities Research Council (grant no. AH/T002859/1).

References

- Abondolo, Daniel. 1998. Finnish. In Daniel Abondolo (ed.), *The Uralic languages*, 149–183. London: Routledge.
- Albright, Adam. 2003. A quantitative study of Spanish paradigm gaps. In Gina Garding & Mimu Tsujimura (eds.), *Proceedings of the West Coast Conference on Formal Linguistics*, Vol. 22, 1–14. Somerville, MA: Cascadilla Press.
- Aller Media ltd. 2014. The Suomi 24 sentences corpus (2016H2) [text corpus]. Kielipankki. Available at: <http://urn.fi/urn:nbn:fi:lb-2017021505>.
- Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412.
- Bader, Markus & Jana Häussler. 2009. Toward a model of grammaticality judgments. *Journal of Linguistics* 45. 1–58.
- Baerman, Matthew. 2008. Historical observations on defectiveness: The first singular non-past. *Russian Linguistics* 32(1). 81–97.
- Baerman, Matthew. 2011. Defectiveness and homophony avoidance. *Journal of Linguistics* 47(1). 1–29.
- Barr, Dale J, Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3). 255–278.
- Bates, Douglas, Martin Mächler, Bolker Ben & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. 1–48.
- Bermel, Neil & Luděk Knittl. 2012. Corpus frequency and acceptability judgments: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory* 8. 241–275.
- Bermel, Neil, Luděk Knittl & Alexandre Nikolaev. Under review. Uncertainty in the production of Czech noun forms.
- Blevins, James P., Petar Milin & Michael Ramscar. 2017. The Zipfian paradigm cell filling problem. In Ferenc Kiefer, James Blevins & Huba Bartos (eds.), *Perspectives on morphological organization*, 139–158. Leiden: Brill.

- Borgatti, Stephen P., Daniel J. Brass Ajay Mehra & Giuseppe Labianca. 2009. Network analysis in the social sciences. *Science* 323. 892–896.
- Bürkner, Paul-Christian. 2017. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software* 80(1). 1–28.
- Bürkner, Paul-Christian. 2018. Advanced Bayesian multilevel modeling with the R package brms. *The R Journal* 10(1). 395–411.
- Bürkner, Paul-Christian. 2021. Bayesian item response modeling in R with brms and Stan. *Journal of Statistical Software* 100(5). 1–54.
- Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.
- Bybee, Joan. 2008. Formal universals as emergent phenomena: The origins of structure preservation. In Jeff Good (ed.), *Linguistic universals and language change*, 108–121. Oxford: Oxford University Press.
- CD-perussanikirja [Basic Finnish CD Dictionary]. 1997. *Publications of the institute for the languages of Finland*, Vol. 94. Helsinki: Institute for the Languages of Finland.
- Chen, Jiahua & Zehua Chen. 2008. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* 95(3). 759–771.
- Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.
- Costantini, Giulio & Marco Perugini. 2020. 17 network analysis for psychological situations. In John F. Rauthmann, Ryne Sherman & David C. Funder (eds.), *The Oxford handbook of psychological situations*, 269–286. New York: Oxford University Press.
- Csardi, Gabor & Tamas Nepusz. 2006. The igraph software package for complex network research. *InterJournal, Complex Systems* 1695(5). 1–9.
- Davies, Mark. 2008. The Corpus of Contemporary American English (COCA). Available at: <https://www.english-corpora.org/coca/>.
- Divjak, Dagmar. 2008. On (in)frequency and (un)acceptability. In Barbara Lewandowska-Tomaszczyk (ed.), *Corpus linguistics, computer tools and applications – State of the art*, 213–233. Frankfurt: Peter Lang.
- Divjak, Dagmar, Natalia Levshina & Jane Klavan. 2016. Cognitive linguistics: Looking back, looking forward. *Cognitive Linguistics* 27(4). 447–463.
- Do, Youngah. 2018. Paradigm uniformity bias in the learning of Korean verbal inflections. *Phonology* 35(4). 547–575.
- Epskamp, Sacha, Angélique O.J. Cramer, Lourens J. Waldorp, Verena D. Schmittmann & Denny Borsboom. 2012. qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software* 48(4). 1–18.
- Epskamp, Sacha, Denny Borsboom & Eiko I. Fried. 2018. Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods* 50(1). 195–212.
- Goldberg, Adele E. 2011. Corpus evidence of the viability of statistical preemption. *Cognitive Linguistics* 22(1). 131–153.
- Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.
- Hakulinen, Auli, Maria Vilkkuna, Riitta Korhonen, Vesa Koivisto, Tarja Riitta Heinonen & Irja Alho. 2004. *Iso suomen kielioppi. SKS:n toimituksia 950*. Helsinki: Suomalaisen Kirjallisuuden Seura.
- Harley, Trevor A. & Siobhan B. G. MacAndrew. 2001. Constraints upon word substitution speech errors. *Journal of Psycholinguistic Research* 30(4). 395–418.

- Harmon, Zara & Vsevolod Kapatsinski. 2017. Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology* 98. 22–44.
- Hedlund, Laura A., Suzanne C.A. Patrik WikmanHut & Alina Leminen. 2021. Neural responses to Finnish inflected forms during overt and covert production: The role of stem frequency and stem allomorphy. *Journal of Neurolinguistics* 57. 100953.
- Hockett, Charles F. 1954. Two models of grammatical description. *Word* 10(2–3). 210–234.
- Janda, A. Laura & M. Francis Tyers. 2021. Less is more: Why all paradigms are defective, and why that is a good thing. *Corpus Linguistics and Linguistic Theory* 17(1). 109–141.
- Janková, Jana & Sara van de Geer. 2018. Inference for high-dimensional graphical models. In Marloes Maathuis, Mathias Drton, Steffen Lauritzen & Martin Wainwright (eds.), *Handbook of graphical models*, 325–350. Boca Raton, Florida, USA: CRC Press.
- Kapatsinski, Vsevolod. 2006. Sound similarity relations in the mental lexicon: Modeling the lexicon as a complex network. *Speech Research Lab Progress Report* 27. 133–152.
- Kapatsinski, Vsevolod. 2010. Frequency of use leads to automaticity of production: Evidence from repair in conversation. *Language and Speech* 53(1). 71–105.
- Kapatsinski, Vsevolod. 2022. Morphology in a parallel, distributed, interactive architecture of language production. *Frontiers in Artificial Intelligence* 5(803259). 1–17.
- Karlsson, Fred. 1983. *Suomen kielen äänne- ja muotorakenne [Structure of Finnish. Phonology and morphology]*. Helsinki: WSOY.
- Karlsson, Fred & Kimmo Koskeniemi. 1985. A process model of morphology and lexicon. *Folia Linguistica* 19(1–2). 207–232.
- Kempen, Gerard & Karin Harbusch. 2008. Comparing linguistic judgments and corpus frequencies as windows on grammatical competence: A study of argument linearization in German clauses. In Anita Steube (ed.), *The discourse potential of underspecified structures*, 179–192. Berlin: Walter de Gruyter.
- Kiparsky, Paul. 1978. Analogical change as a problem for linguistic theory. *Studies in the Linguistic Sciences* 8(2). 77–96.
- Kirjavainen, Minna, Alexandre Nikolaev & Evan Kidd. 2012. The effect of frequency and phonological neighbourhood density on the acquisition of past tense verbs by Finnish children. *Cognitive Linguistics* 23(2). 273–315.
- Kittredge, Audrey K., Gary S. Dell, Jay Verkuilen & Myrna F. Schwartz. 2008. Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology* 25(4). 463–492.
- Laine, Matti & Patrik Virtanen. 1999. *WordMill lexical search program*. Turku: Centre for Cognitive Neuroscience, University of Turku.
- Langacker, Ronald W. 2016. Working toward a synthesis. *Cognitive Linguistics* 27(4). 465–477.
- Levenshtein, Vladimir I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics - Doklady* 10(8). 707–710.
- Lieberman, Erez, Jean-Baptiste Michel, Joe Jackson, Tina Tang & Martin A. Nowak. 2007. Quantifying the evolutionary dynamics of language. *Nature* 449(7163). 713–716.
- Milin, Petar, Dagmar Divjak, Strahinja Dimitrijević & Harald R. Baayen. 2016. Towards cognitively plausible data science in language research. *Cognitive Linguistics* 27(4). 507–526.
- Newman, Mark E. J. 2010. *Networks: An introduction*. Oxford, UK: Oxford University Press.
- Nikolaev, Alexandre & Neil Bermel. Under review. Studying negative evidence in Finnish language corpora.

- Nikolaev, Alexandre, Pääkkönen Ari, Jussi Niemi, Mikko Nissi, Eini Niskanen, Mervi Könönen, Esa Mervaala & Hilikka Soininen. 2014. Behavioural and ERP effects of paradigm complexity on visual word recognition. *Language, Cognition and Neuroscience* 29(10). 1295–1310.
- Nikolaev, Alexandre, Minna Lehtonen, Eve Higby, JungMoon Hyun & Sameer Ashaie. 2018. A facilitatory effect of rich stem allomorphy but not inflectional productivity on single-word recognition. *Applied Psycholinguistics* 39(6). 1221–1238.
- Nikolaev, Alexandre, Sameer Ashaie, Merja Hallikainen, Tuomo Hänninen, Eve Higby, JungMoon Hyun, Minna Lehtonen & Hilikka Soininen. 2019. Effects of morphological family on word recognition in normal aging, mild cognitive impairment, and Alzheimer's disease. *Cortex* 116. 91–103.
- Nikolaev, Alexandre, Eve Higby, JungMoon Hyun, Minna Lehtonen, Sameer Ashaie, Merja Hallikainen, Tuomo Hänninen & Hilikka Soininen. 2020. Production of inflected novel words in older adults with and without dementia. *Cognitive Science* 44(8). e12879.
- Pertsova, Katya. 2016. Transderivational relations and paradigm gaps in Russian verbs. *Glossa: A Journal of General Linguistics* 1(1). 1–34.
- R Core Team. 2020. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: <https://www.R-project.org/>.
- Sims, Andrea. D. 2015. *Inflectional defectiveness*. Cambridge University Press.
- Sims-Williams, Helen & Hans-Olav Enger. 2021. The loss of inflection as grammar complication: Evidence from Mainland Scandinavian. *Diachronica* 38(1). 111–150.
- Smolek, Amy & Vsevolod Kapatsinski. 2018. What happens to large changes? Saltation produces well-liked outputs that are hard to generate. *Laboratory Phonology* 9(1). 1–21.
- Stave, Matthew, Smolek Amy & Vsevolod Kapatsinski. 2013. Inductive bias against stem changes as perseveration: Experimental evidence for an articulatory approach to output-output faithfulness. In *Proceedings of the 35th annual meeting of the cognitive science society*, 3454–3459. Austin: Cognitive Science Society.
- Suomen kielen perussanakirja* [Basic dictionary of Finnish]. 1990–1994. Helsinki: Edita Oyj.
- Theakston, Anna L. 2004. The role of entrenchment in children's and adults' performance on grammaticality judgment tasks. *Cognitive Development* 19(1). 15–34.
- Thornton, Anna M. 2011. Overabundance (multiple forms realizing the same cell): A noncanonical phenomenon in Italian verb morphology. In Martin Maiden, John Charles Smith, Maria Goldbach & Marc-Olivier Hinzelin (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 358–381. Oxford: Oxford University Press.
- Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B* 58(1). 267–288.
- van der Loo, Mark P. J. 2014. The stringdist package for approximate string matching. *The R Journal* 6(1). 111–122.
- Vea, Joakim & Christer Johansson. 2020. Priming paradigmatic gaps. In Vincent Torrens (ed.), *Typical and impaired processing in morphosyntax*, 117–133. Amsterdam: John Benjamins.
- Wagner, Robert A. & Roy. Lowrance. 1975. An extension of the string-to-string correction problem. *Journal of the Association of Computing Machinery* 22(2). 177–183.
- Yan, Erjia & Ying Ding. 2009. Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology* 60(10). 2107–2118.

- Yin, Sora Heng & James White. 2018. Neutralization and homophony avoidance in phonological learning. *Cognition* 179. 89–101.
- Zipf, George Kingsley. 1949. *Human behavior and the principle of least effort: An introduction to human ecology*. Boston: Addison-Wesley.

Supplementary Material: The online version of this article offers supplementary material (<https://doi.org/10.1515/cog-2021-0041>).