Article:

Bermel, N. orcid.org/0000-0002-1663-9322, Knittl, L. and Nikolaev, A. (2023) Uncertainty in the production of Czech noun and verb forms. Word Structure, 16 (2-3). pp. 258-283. ISSN 1750-1245

https://doi.org/10.3366/word.2023.0231

Uncertainty in the production of Czech noun and verb forms[1]

Neil Bermel (University of Sheffield)
Luděk Knittl (University of Sheffield)
Alexandre Nikolaev (University of Eastern Finland)

Abstract

We examine the reactions of Czech native speakers to cues asking them to supply inflectional forms of nouns and verbs that are either canonical (non-variant), overabundant, or supposedly defective, to see what distinguishing characteristics these three conditions have for production. We find that respondents handle defective material differently from other conditions, producing different sorts of forms at different frequencies, and taking significantly longer to do so. Overabundant cells pattern at the individual level like canonical inflectional cells, but collectively display a significantly more varied and less focused spread of forms produced than our canonical cells. The individual dimension of uncertainty in production is thus limited to defective cells, but the collective dimension of uncertainty is evident between all three conditions.

Keywords

overabundance, defectivity, reaction times, inflection, Czech

1. Introduction

An organisational principle of language is that a correspondence existing between form and meaning is supposed to be systematic (Booij 2009). Finding multiple word forms that refer to the same content, we expect each to serve some distinct function; and identifying a function we could make use of with a particular lexeme, we expect to find a single form representing it.[2] Lexemes that have *defective* or *overabundant* cells in an inflectional paradigm fail to respect this principle. If we take the concept of canonicity (Corbett 2005) as reflecting an idealized inflectional paradigm in which each function maps to a single, unique form, then defective cells deviate from it by lacking a usable form, while overabundant cells deviate from it by having multiple forms that can serve a single function. An example of a defective inflectional paradigm in English would be *stride* (simple past *strode*, participle unclear), while an overabundant paradigm in English would be *strive* (simple past *strove,* participle *strived* or *striven).*

Studies on defective cells (Sims 2009: 3, Maiden & O'Neill 2010: 106) have pointed out that the gap in an inflectional paradigm frequently coincides with the coexistence of variant possibilities, often when multiple plausible stems are involved (do we generate the participle of *stride* based on a pattern like *ride, rode, ridden* or based on one like *bide, bode, bided*?), although it is widely acknowledged that this coexistence is not sufficient to motivate the

[1] This research was supported by grant AH/T002859/1 from the UK Arts and Humanities Research Council as part of the Feast and Famine project (www.sheffield.ac.uk/feastandfamine).
[2] In historical studies, the corollary is that observed variation is over time supposedly reduced and eliminated by this drive towards *isomorphism*; for a critical discussion of this, see De Smet et al. (2018: 198–201).

appearance of a gap.[3] A second frequently cited motivation for defective cells is that the form involved has already been "co-opted" for use in a different context. For example, Russian *pobedit'* 'vanquish/win' has most plausibly 1 sg. *pobežu* 'I will vanquish', but that form would more logically be a non-standard 1 sg. of the more frequent verb *pobežat'* 'run', cf. similar examples in Moskvin (2015: 211); Baerman (2011) admits a role for homophony in causing avoidance but is skeptical that any cross-linguistic generalizations about its operationalization can be made.[4]

Overabundant cells as per Thornton (2011, 2012) can result from the existence of variant plausible stems (*one die/dice, two dice; one octopus, two octopi/octopuses)*, as well as from the availability of multiple plausible exponents (*James'/James's car*); the latter tends not to be listed as a potential motivator for gaps.

Fig. 1 – Co-occurring factors with defective and overabundant cells

<table>
<tr><td>Co-option of form</td><td>Multiple plausible stems</td><td>Multiple plausible exponents</td></tr>
<tr><td colspan="2">DEFECTIVE CELL</td><td></td></tr>
<tr><td></td><td colspan="2">OVERABUNDANT CELL</td></tr>
</table>

The three co-occurring factors (Fig. 1) present the producer with *uncertainty*: a lack of ability to "automate" a choice. Faced with a diversity of possibilities, the speaker or writer must select one or find another way to convey the information that disposes with the uncertainty.

In a scenario with *multiple plausible stems*, then, three resolutions are possible: (a) the uncertainty resolves consistently in favour of one form (non-variance); (b) the uncertainty persists, and two or more forms are found and accepted (overabundance); (c) the use of any of the forms is avoided (defectivity). The multiple plausible stems condition is thus the best possible test bed for comparing these three potential outcomes. Our research questions were: Do respondents treat these cells the way we might expect from corpus data and information in handbooks? Can we detect anything in their responses that is characteristic of uncertainty? In Nikolaev & Bermel (2022) we operationalized uncertainty both as a *collective* phenomenon (for a cohort of speakers) and as an *individual* phenomenon (for each given speaker) and applied this to evaluate the possibility that cells in a Finnish verbal paradigm with multiple stems are either defective or non-defective. We found that respondents could adequately evaluate uncertainty using several metrics (perceived frequency, contemporaneousness, and acceptability), and that the frequency of both the lemma and the specific form in question played a central role in their evaluations. Respondents reacted to this uncertainty in a gradated fashion, sometimes avoiding filling a defective slot, but more frequently using periphrasis, etymologically related similar lexemes, or novel forms to fill the gap. The current contribution expands the scope of this enquiry to include overabundance as

[3] This account assumes a whole-form view of morphology; an account tied to morphophonology would in some instances lean on ways of resolving potential clashes in realization rules, as per Gouskova & Becker (2013); see Sims (2017: 492–496) for an exploration of these accounts. Scholars have cast doubt on the motivation of realization-rule accounts: see *inter alia* Sims (2009), Daland, Sims & Pierrehumbert (2007), Baerman (2008).

[4] Based on overall lexeme frequency, this is not an exceedingly strong example. The Russian National Corpus records 24,017 occurrences (58.2%) of *pobežat'* and 17,223 (41.8%) of *pobedit'* (www.ruscorpora.ru, 13 November 2021). The fact that the dispreferred reading (1 sg. *pobedit'*) contains a stem differing by way of consonant mutation and the preferred reading (1 sg. *pobežat'*) does not is a valid observation, but not an explanation, especially as the "standard" 1 sg. of *pobežat'* is in fact the irregular form *pobegu.*

a third possible formal instantiation of uncertainty, and to test its relationship to the other two possible outcomes (defective and non-variant).

2. Corpus and handbook data (selection/identification of lexemes)

Our data come from Czech, a language with rich inflectional morphology: nouns have six syntactic cases and a vocative form, and two numbers, while verbs have three persons and two numbers across two synthetic tenses and several analytic tenses and moods. Standard Czech is a supranational code that is used primarily in writing and formal speech and is acquired at school; the subject of extensive cultivation activity for 250 years, it has over the last century admitted many variant forms from common spoken varieties, mostly without losing those perceived as characteristic for the standard. In part because of these developments, standard Czech has many overabundant paradigm cells. At the same time, the "engineered" nature of the standard means that *naturalness* is often trumped by potential *usefulness;* a form can be introduced to fill a perceived gap on structural grounds without native speakers (NS) protesting its novelty.[5] Consequently, handbooks rarely allude to defective paradigm cells, preferring to offer a possibility rather than a gap.

Our first goal was thus to identify elusive defective cells. Fortunately, stem alternations occur in reliable places in Czech. For nouns, they occur where a stem-final consonant cluster meets a vocalic inflectional suffix vs. where it meets a zero inflectional suffix, causing an inserted epenthetic vowel before the stem-final consonant (*kalkulačk-a* 'calculator-NOM.SG' vs. *kalkulaček-Ø* 'calculators.GEN.PL'). For verbs, the so-called "class 1" conjugation (whose non-past inflectional endings are built on the theme vowel -e-) shows stem alternations across an infinitive stem and potentially up to three further stems: a past stem, a non-past stem and a passive stem (cf. *ps-á-t* 'write-INF' vs. *píš-u* 'write-1.SG'; *nadch-n-ou-t* 'enthuse-INF' vs. *nadš-en-Ø* 'enthuse-PASS-MASC.SG'). We relied on handbooks, searches in electronic versions of the two major dictionaries of Czech, and occasional mentions in the scholarly literature, and then cross-referenced our findings with data from large-scale corpora of written Czech.[6]

Grammar handbooks consulted (*Mluvnice současné češtiny; Internetová jazyková příručka)* made occasional mention of nouns and verbs where a form or class of forms was not recommended, and online dictionary searches (*Příručný slovník jazyka českého; Slovník spisovného jazyka českého; Internetová jazyková příručka*) yielded items where a form was marked as 'rare' or 'not used'. This yielded a list of around 50 nouns and verbs, focusing on (1) nouns where the genitive plural form is said to be absent; (2) nouns said to have no oblique case forms; (3) verbs said to have no forms in the non-past tense; (4) transitive verbs said to have no passive participle.

We verified our findings using csTenTen17 (Suchomel 2018) and the Czech National Corpus, primarily the SYN2015 100-million-token representative corpus (Křen et al. 2016), checking all the lexemes in the sample and using wildcard searches to add others with similar features (small subclasses of verbs with multiple stems; nouns whose final consonant clusters show low neighborhood density). Many of the nouns and some of the verbs searched had

[5] Unless it is a slot or feature that has over time become thematized in language planning discourse; for examples, see Bermel (2007: 74-75; 2014: 30–32).

[6] Corpora of spoken Czech contain < 6m tokens, which is only enough to explore the most frequent lexemes for variation in their morphological forms. Subtitle corpora, which are often used as a proxy for spoken language corpora, are in the case of Czech translation-focused and not yet lemmatized or tagged.

very low frequencies and were from an informal register, making TenTen's size and composition suitable for this task, but its tagging and lemmatization meant much manual clean-up and verification was needed; in the case of higher-frequency items this quickly became unwieldy. GramatiKat, a corpus tool that allows searching of nouns by percentage of case and number forms (Kováříková & Kovářík 2021; Kováříková 2021), made it easy to identify medium- to high-frequency nouns in the Czech National Corpus where the relevant cells were suspiciously underweighted; we looked at all such items for inclusion. All items were then checked manually in the 100m-token SYN2015 corpus.

To summarize, defective items or classes were identified in reference works, verified in corpora and vetted by a native speaker. At the end of this process, we had a list of 17 nouns and 11 verbs with defective cells.

We repeated this process with minor variations to identify similarly positioned cells that were overabundant. We therefore looked at: genitive plural noun forms, non-past verb stems and passive verb stems, where multiple stems thus gave rise to variant forms for a given slot. This approach meant that some variation was not included in our lexical stock. For example, the noun *ambice* 'ambition' has alternate gen. pl. forms *ambic/ambicí*, but there is no possible stem variation evident here, so it was excluded. The goal was to avoid confounding differences between defective and non-defective paradigms with differences between single-stem and multiple-stem lexemes. To qualify, a slot had to be mentioned in a major handbook and attested in the Czech National Corpus with evidenced competition between two forms (i.e., at least 10% of each form).[7]

As a control, a set of lexemes where variation is not expected was drawn up on the same principles, using data from handbooks and checking it against the Czech National Corpus. Here too, only lexemes where stem variation was plausible or actual were used. We call these our 'non-variant' lexemes.[8] For example, *šelma* 'wild animal' has the stems *šelm-, šelem-,* but only the latter appears in the gen. pl., while the former is used in all other cases, which have a vowel in desinence-initial position. Conversely, *tango* 'tango' has the plausible stems *tang-, taneg-,* but only the former is attested. Although in the first instance both plausible stems appear and in the second instance only one plausible stem appears, neither lexeme displays variation in the corpus and thus both exemplify the non-variant condition.

An observation about the corpus data, to which we will return later in the analysis, is that our defective nouns and our verbs with defective non-past stems are typically of low or very low frequency, while our transitive verbs with defective passive participle stems are typically of high frequency. While we can posit uncertainty arising from unfamiliarity or rarity in the first two instances, we cannot posit it in the final instance. In any event, we used linear mixed-

[7] This introduced one issue to which we will return in the discussion in section 5: while many of the defective nouns were of very low frequency, we were unable to use extremely low-frequency lexemes to identify overabundant cells, because there needed to be enough forms in the corpus to make a reasoned judgment that overabundance had been found.

[8] For example, vowel length in Czech is considered phonologically distinctive. A long final stem vowel in forms with a syllabic inflectional ending often corresponds to a short final stem vowel before a zero ending occurring in the gen. pl. (e.g., *kráva* 'cow' has gen. pl. *krav*), and because these are phonologically distinct vowels, it yields two attested stems (*kráv-/krav-*). The opposite effect – a short vowel in forms with a syllabic inflectional ending corresponding to a long vowel before a zero ending in the gen. pl. – is not typically found, so *hala* 'hall' does not have a plausible second stem **hál-* and such lexemes were not evaluated for inclusion.

effects models to explain inflectional choices and reaction times, and accounted for frequency by adding it to the model as a potential explanatory variable.

In asking whether multiple-stem nouns and verbs we chose produce measurable uncertainty for respondents, we designed an experiment to evaluate the following hypotheses, choosing to operationalise *uncertainty* through collective variety and regularity of choice (as in our previous study) and adding *reaction time* as a further indicator:

- **H1**: Czech native speakers produce a greater variety of forms for D (defective) and OA (overabundant) cells than for non-variant (N) cells.
- **H2**: Czech NSs produce the 'most regular' choices less often for D and OA cells than for N cells.
- **H3**: Czech NSs are slower at producing forms to fill D and OA cells than they are for N cells.

We predicted in each instance that the values for D, OA and N would differ in predictable ways. For each, D < OA < N, where '<' can mean 'less uniformity', 'less often' or 'less decisiveness (longer reaction time)', whereas for the null hypotheses, there would be no distinguishable differences between D, OA and N cells.

3. Materials and methods

3.1. Participants

A total of 144 participants (native Czech speakers) took part in this study. Sixty respondents were randomly assigned to the noun task and 84 to the verb task (the characteristics of our cohorts are presented in Table 1). Upon completion of the task, the participants were offered the opportunity to claim a book token worth 100Kč (about £3) by providing their email address. Respondents were recruited through personal contacts and social media.

3.2. Material and procedure

The experiment consisted of two tasks, one on nominal lexemes and one on verbal lexemes. The task was administered using Gorilla, an online data collection tool.[9] Online data collection was a necessity in this period, although in any event studies have shown that online studies give adequate results, except for a few task types needing extremely brief presentation intervals, fine color discrimination or carefully calibrated audio (Woods et al. 2015). The sort between verb and noun task was random, although we set Gorilla to skew towards more respondents for verbs than nouns: as the verb survey was shorter, we needed more respondents to make sure it had adequate predictive power.

The first page of the survey asked for basic personal data (age, education, region of origin) and respondents were subsequently sent to either the verb or the noun task. The task took

[9] Ethical approval for this experiment was sought and received through the University of Sheffield's ethics review process. The original plan had been to collect data in person on a single computer, for consistency's sake, but due to the pandemic, in-person data collection was not possible. The use of remote collection introduces some difficulties, such as possible differences in ease of input between devices, more possibility of misunderstanding the task or abandoning it partway through, etc. This was to some extent compensated for by increasing the survey size and careful attention to how the answers provided were included. The use of Gorilla, which has reliable reaction time monitoring, was a further help in this regard.

roughly 10 minutes to complete, and each participant completed only one task (either noun or verb task).

The nominal lexeme task tested 51 lexemes, evenly split between those with D, OA and N cells as described above. The verbal lexeme task tested 33 items, again evenly split between these lexeme types. On an introductory screen, the respondents saw an example demonstrating what to do. Within the task, on each screen, the respondent saw one sentence with an item highlighted, and a second sentence with a gap. Below the sentences was a box in which they typed in the answers and then pressed a button to proceed to the next screen. Short fixation screens preceded each trigger sentence. The order of screens was randomized to avoid order effects.

For the noun task, the first sentence contained a noun in a direct case (nominative or accusative), while the second sentence required a form where an alternate stem was possible, typically the genitive plural (example 1):

1. Na okruhu je **zácpa**. V dopravní špičce je vždycky mnoho ……
   ‘There is a **traffic jam** on the ring road. At rush hour there are always many …... GEN.PL.’
   zácp-a 'traffic jam-NOM.SG'

For the verb task, the first sentence contained a finite verb form in the past tense, while the second sentence required a form in the non-past tense, which can be built on a different stem (example 2); alternatively, the first sentence contained a verb in the active voice, while the second sentence required a passive participle, which can be built on a different stem (example 3).[10]

2. Vykonal to, co si **předsevzal**. A zítra si zase ……. něco nového.
   ‘He did what he **had resolved.** And tomorrow he …….NPST-3.SG something else.’
   předsevza-l-Ø 'resolve-PST.M.SG'

3. Vyjednavači **našli** uspokojivé řešení. Uspokojivé řešení bylo …. našimi vyjednavači.
   ‘The negotiators **found** a satisfactory solution. A satisfactory solution was ……PASS-N.SG by our negotiators.’
   naš-l-i 'find-PST-M.PL'

Care was taken to ensure that, if the instructions were followed, the form needed would be clear without further directions.[11]

3.3. Data analysis

[10] There was no hypothesis concerning the difference between the two sorts of verb tasks; the existence of two different types was solely down to the scarcity of defective items, necessitating the inclusion of two such types in one questionnaire. However, we did then include the task type as an explanatory variable in the analysis (see 4.3).

[11] For nouns, this meant that the context had a clear marker of case, such as a preposition or quantifier, and where needed, a clear marker of number, such as a declined adjective. For verbs, we indicated tense with adverbs such as *zítra* ‘tomorrow’ and indicated passive voice through a copular verb and the demotion of the agent to the instrumental case.

Before analysis, the data were examined and adjusted to avoid misleading results. These were largely of two sorts: typing inconsistencies and timing issues.

Typing inconsistencies were adjusted as follows: (1) distinctions between upper-case and lower-case letters were normalized to all minuscule, so that e.g. *jablek, JABLEK* and *Jablek* 'applesGEN.PL' were not counted as three different forms just because some users' devices had caps lock on or automatically capitalized first letters of words; (2) although respondents had been directed to use diacritics as is standard in Czech, some nonetheless left them out consistently, which is a feature of informal electronic communication. For these respondents, we inserted accent marks as found in the 'standard' forms of the words.[12] These measures ensure that purely formal, visual variation does not skew our results.

Timing issues are a hazard of unmonitored surveys. While Gorilla's measurement of timings has been found to be relatively accurate (Anwyl-Irvine et al. 2020), respondents may stop mid-survey for various reasons, in which case either of the reaction time (RT) measurements (initial RT, the point at which the respondent starts typing; or final RT, the point at which the respondent finishes typing) can be misleadingly long. To forestall this, we trimmed all times pre-analysis to a maximum of 15 seconds for initial RTs and 20 seconds for final RTs, which was 1.5x the interquartile range.

All the statistical analysis were done in R (R Core Team, 2021). We analyzed the Reaction time (RT) data using a mixed effects model (Bates, Mächler, Bolker, & Walker, 2015; Kuznetsova, Brockhoff, & Christensen, 2013), which takes account of the fact that we have variables of both interval (e.g., time) and nominal (e.g., type of condition) type. Our models included participants, items, and trial numbers as random effect variables. Including random effect variables in the models accounts for repeated measurements (we had more than one answer/measurement from each participant, and we had more than one participant responding to each item). Ignoring repeated measurements (treating them as independent data points) would inflate the rate of false positive findings. Variation in these is considered to be part of the overall normal dispersal of answers in the survey, as participants have individual differences, items may be harder or easier, etc. The fixed-effect factors are those that will influence the answers systematically. We also added possible interactions between word type and the task type into the models. The significant interaction was visualized using the "sjPlot" package (Lüdecke, 2017). Log-transformed reaction times were the response variable (the one that we explain / predict with random and fixed-effect variables). Logarithmic transformation of RTs typically makes them more normally distributed. Following Baayen and Milin (2010), to improve the model and remove the influence of possible outliers (in this case, excessively long reaction times), we excluded data points with absolute standardized residuals exceeding 2.5 standard deviations. We also added by-participant random slopes for some predictors (e.g., Length) into the models, which help to account for individual differences in the effects of our predictor variables; however, if the model comparison showed that they did not improve the fit of the model, we took them out.

The models where the dependent variable had only two values (e.g., expected inflection vs. other inflection) were done using generalized linear mixed-effects models. As with the models for RTs, each new explanatory variable or term of interaction was compared to the

[12] However, where a user had otherwise used diacritics, we treated their occasional absence as salient and did not add them, and we did not correct otherwise obvious typos, e.g., *rzn* for *zrn* 'grain.GEN.PL', as there were no consistent principles on which to do so.

model without it (using the function *anova*). We also compared Akaike Information Criteria (AIC) indices of the models; these can be seen as a measure of "efficiency" that weigh the accuracy of a model against the amount of information used to predict outcomes. A model with a lower AIC index (meaning better fit) was preferred.

Nominal scale predictors – for example, "word type" as D, OA or N – in the mixed-effects models only indicate whether one condition (e.g., N) has a different effect on the dependent variable (e.g., RT) compared to the other two conditions (D and OA). However, the model does not tell us if there is a difference between the latter two (D and OA) in their effect on the dependent variable. Of course, it is possible to change the reference (e.g., to D) and re-run the model in order to see if the OA condition has a different effect than the D condition, however, a more reliable result can be achieved using post-hoc analyses for the pairwise comparisons between D, OA, and N types: we conducted these using simultaneous tests for general linear hypotheses with adjusted p-values.

The 100m-token SYN2015 corpus was used to obtain corpus frequencies for the experiment items. We were interested in whether the existence of phonemically closely related words – "neighbours" – had an effect on production; therefore, we made use of two different measures for this, one weighted towards similarities at the end of the word, where inflectional material is placed in Czech, and another that measures difference anywhere in the word. Neighbourhood density and a Hamming distance of 1 for the experiment items were calculated by counting the number of words with the same length but differing in the initial letter (neighborhood density) or in any letter anywhere in the word (HD1) using the lemmas from the Czech Internet Language Handbook, which is an online dictionary based on the standard Dictionary of Literary Czech.

4. Results

The characteristics of our cohorts are presented in Table 1.

Table 1. Characteristics of respondents to both tasks

	Noun task	Verb task
Gender		
Male	20	26
Female	40	57
Other	0	1
Age		
Mean (sd)	44 (17)	43 (17)
Education		
Primary	2	5
Secondary	20	30
Tertiary	38	49
Region		
Karlovarský	0	0
Ústecký	2	3
Liberecký	4	6
Plzeňský	2	2
Jihočeský	0	4
Středočeský	3	4
Praha	14	18
Královéhradecký	0	0
Pardubický	3	8
Vysočina	1	3
Jihomoravský	6	10
Zlínský	4	4
Olomoucký	9	8
Moravskoslezský	12	14

Women are thus overrepresented in the survey; this is a common feature noticed in many such surveys.[13] The average age in the noun task was 44 years. One participant did not enter his/her age, and thus we substituted the mean value. The average age in the verb task was 43 years. Three participants did not enter their age, and these were substituted for by the mean value.

4.1 Collective uncertainty (number and popularity of answers)

Our first hypothesis concerned the number of different results produced. We hypothesized that respondents would produce the most variant forms for D items, fewer for OA items, and fewest for N items. A corollary to this is the extent to which answers converge on a single, most-used variant.

For nouns, the number of forms generated per target cell are given in Table 2. A parametric test (Tukey's 'Honest Significant Difference') was used to test for differences; while D cells generate significantly more forms than OA (p-value < 0.001) or N cells (p-value < 0.001), the difference between OA and N cells is not significant. (A non-parametric equivalent of Tukey's test gave similar results.)

Table 2. Number of forms generated by respondents for nouns

[13] A series of studies on morphology we ran over the last decade have consistently drawn a preponderance of female respondents (inter alia Bermel, Knittl & Russell, 2018: 23; Bermel, Knittl & Russell 2015: 291; Bermel & Knittl, 2012: 108), using different recruitment methods and survey tools. Gender was entered as a variable in the analyses but was not significant in any of them. In a recent study (Nikolaev & Bermel, 2022), an initial sample was also skewed towards women. We subsequently recruited more men to achieve a roughly 50:50 balance. However, neither the pilot analysis with skewed gender data, nor the final analysis with balanced gender data, showed significant differences between genders.

Defective cells		Overabundant cells		Non-variant cells	
mamča	9	slívu	7	tango	6
pestřec	9	křovisko	5	objížďka	5
pivčo	9	rolba	5	pomsta	5
jařmo	8	šťáva	5	hra	4
kostřec	8	boudu	4	jablko	4
msta	8	jachtu	4	šelmu	4
ostropestřec	7	bulva	3	tundra	4
vindru	7	hledisko	3	zrno	4
zácpa	7	mísa	3	bříza	3
babča	6	šichta	3	káva	3
přerva	5	strouha	3	plátno	3
čtvrtka	4	váhy	3	sklo	3
limča	4	víra	3	vrána	3
ségru	4	blána	2	houba	2
parka	3	bránu	2	kostra	2
průrva	3	hlína	2	moucha	2
šmouha	3	víla	2	sprcha	2
Total no. forms	104		59		59
Mean	6.12		3.47		3.47
Median	7		3		3
SD	2.19		1.33		1.14

For verbs, the number of forms generated per target cell are given in Table 3. A parametric test (Tukey's 'Honest Significant Difference') was used to test for differences; while D cells generate significantly more forms than OA (p-value < 0.001) or N cells (p-value < 0.001), the difference between OA and N cells is not significant. (A non-parametric test gave similar results.)

Table 3. Number of forms generated by respondents for verbs

Defective cells		Overabundant cells		Non-variant cells	
měli	32	sypali	10	přeskočili	6
předsezval	23	tiskli	10	vypověděli	6
klál	21	vyvlékl	10	dali	5
nevíme	19	odemkl	8	vymazal	5
zapovědělo	17	dotkl	7	objevili	4
jala	15	otiskli	6	sejmuli	4
oběšli	15	nekousala	5	zvládli	4
přešli	14	oblékl	5	převzali	3
našli	9	vyřkl	5	zapomněli	3
znali	9	zatkli	4	bál	2
minuli	4	odkašlal	3	rozhodly	2
Total no. forms	178		73		44
Mean	16.18		6.64		4
Median	15		6		4
SD	7.67		2.54		1.41

So far we have looked at the variety of forms produced. Next, we take the *convergence between respondents on the most popular form* as a measure of collective uncertainty. We looked at results where obvious typos (missing letters added, adjacent letters substituted, etc.) had been corrected to an expected form. In charts 1 and 2, we can see three distinct profiles for both nouns and verbs. The most popular non-variant form always represents over 90% of answers and is in every instance the answer we would have predicted.[14] The most popular defective form rarely represents over 75% of answers, with the interquartile range of the most popular form constituting from 30–60% of answers. The most popular overabundant form is somewhere in between these two extremes: its interquartile range overlaps with those of defective answers but not with those of non-variant ones.

Chart 1. Convergence on most popular noun answer (no. of respondents)

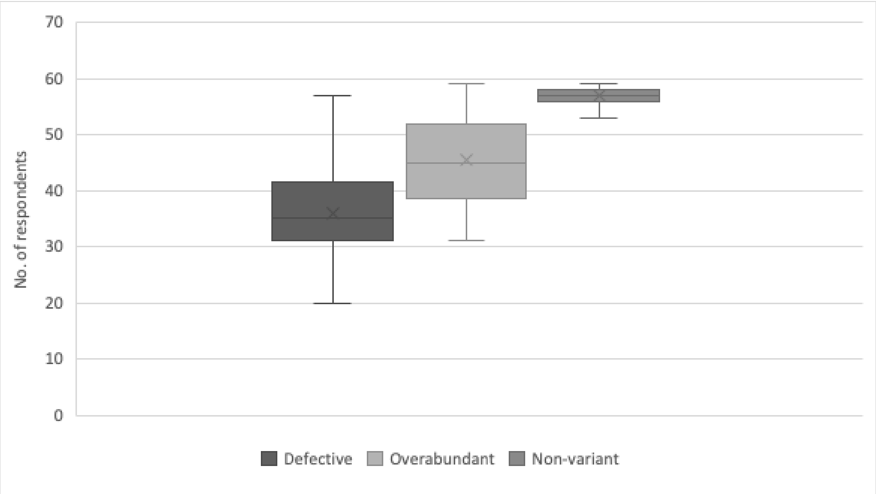
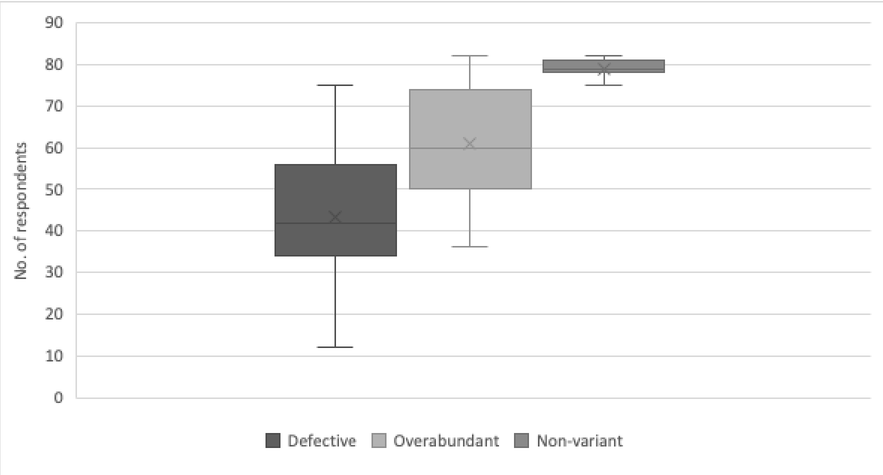


Chart 2. Convergence on most popular verb answer (no. of respondents)



[14] There were in almost all cases some deviations, represented either by choosing a different word that the respondent thought was superior in the context, or by a mistakenly included incomplete survey (sometimes the final answer was not recorded).

A parametric test (Tukey's 'Honest Significant Difference') was applied to these three categories and the differences were found to be statistically significant for nouns (N~D, p-value < 0.001; OA~D, p-value = 0.001; OA~N, p-value < 0.001) as well as for verbs (N~D, p-value < 0.001; OA~D, p-value = 0.008; OA~N, p-value < 0.008).

4.2. Production of forms and contributing factors

4.2.1 Nouns. Expected vs. other forms.

Our second hypothesis concerned the extent to which Czechs would produce the most regular (most easily morphologically predictable) forms. For nouns, we ran a generalised mixed-effect model for inflectional choices as the dependent (response) variable that were coded either 0 (morphologically closest or expected) or 1 (other than expected forms). The model is presented in Table 4. The reference category for Lexeme type (D, OA, N) is N (non-variant cells) and that is why it is not visible in the model. The reference category for the categorical variable Education (primary, secondary, tertiary) is secondary. Since each noun had two (rather than one) "expected" (morphologically closest) forms, one more common and another less common, their logarithmically transformed corpus frequencies (lemma fq) were also added to the model.

Table 4. Inflectional choices model for nouns

Fixed effects	Estimate	Std.Error	*z*-value	Pr(>|z|)
(Intercept)	-2.548	0.843	-3.024	0.002
Lexeme D	1.958	0.751	2.609	0.009
Lexeme OA	0.148	0.639	0.232	0.817
Tertiary education	-1.631	0.480	-3.395	<0.001
Primary education	-0.095	1.241	-0.076	0.939
Expected form fq 2	-0.418	0.151	-2.781	0.005
Expected form fq 1	-0.309	0.161	-1.922	0.055
Random effects				
Groups	Name	Variance	Std.Dev.	
Participant	(Intercept)	2.477	1.574	
Lexeme	(Intercept)	2.136	1.462	
Trial	(Intercept)	0.0218	0.148	
Number of observations: 3037				
Participants: 60				
Lexemes: 51				
Trials: 51				

Participants produced fewer expected variants for D nouns and more expected variants for OA and N nouns. The difference between D and OA is, however, not that large (and not significant). The difference between D and N is significant, while the difference between the OA and N items is not significant.

Lemma frequency did not appear as a significant effect (cf. the RT model below); instead, it turns out that corpus frequencies of the less-common forms (Expected form fq 2) are a significant predictor of participants' inflectional choices. The frequency of individual potential forms thus plays a role, while the overall lemma frequency does not.

Education plays a role, in that those with higher (tertiary) education produced more expected forms than people with secondary education (the primary-educated group did not show a significant difference).

Other variables such as gender, age, and neighbourhood density were not significant and therefore were excluded from the model.

4.2.2 Nouns. Expected forms only (those with a stem change vs. those without a stem change)

We replicated these analyses, including only *expected* inflected forms (in other words, removing all non-expected inflectional choices, which constituted about 11% of responses), and the results can be found in Table 5. Expected inflected forms give us insight into the role of competition between two stems.

Table 5. Inflectional choices model for two expected variant forms (nouns)

Fixed effects	Estimate	Std.Error	z-value	Pr(>|z|)
(Intercept)	-1.988	1.600	-1.242	0.214
Lexeme D	0.933	1.600	0.583	0.560
Lexeme OA	-1.899	1.261	-1.506	0.132
Tertiary education	0.526	0.220	2.383	0.017
Primary education	0.324	0.581	0.557	0.577
Expected form fq 2	-1.359	0.309	4.398	<0.001
Expected form fq 1	-0.514	0.314	-1.635	0.102
Random effects				
Groups	Name	Variance	Std.Dev.	
Participant	(Intercept)	0.366	0.605	
Lexeme	(Intercept)	10.327	3.213	
Trial	(Intercept)	0	0	
Number of observations: 2708				
Participants: 60				
Lexemes: 51				
Trials: 51				

Once non-expected forms are removed, Lexeme type (D, OA, or N) is no longer a significant predictor of exactly which expected form is produced. Higher (tertiary) education predicts more Expected 2 inflected forms than Expected 1; this means that respondents with a degree favoured forms with stem changes over those with no stem changes. Corpus frequencies of Expected form 2, which contains stem changes, predict results better than those of Expected form 1, which does not. This suggests that, when we consider only the choice between the two most expected forms, it is the frequency of the more “distant” form that determines people’s choices, rather than the frequency of the more easily derivable form.

4.2.3 Verbs. Expected vs. other forms.

Turning to verbs, we find a more complex situation, as there are frequently two or more stems and several options for inflectional endings. In this model we coded all forms on the two most expected stems as 0 and other (unexpected or novel) forms as 1. Three predictors turned out to be significant: Lexeme type, Education, and Length of string. The results are given in Table 6.

Table 6. Inflectional choices model for verbs

Fixed effects	Estimate	Std.Error	*z*-value	Pr(>|z|)
(Intercept)	-7.543	0.863	-8.744	<0.001
Lexeme D	6.278	0.855	7.343	<0.001
Lexeme OA	1.975	0.846	2.334	0.020
Tertiary education	-0.706	0.230	-3.077	0.002
Primary education	-0.432	0.475	-0.911	0.363
Length	0.372	0.066	5.624	<0.001
Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	2.183	1.477	
	Length	0.037	0.191	-0.88
Lexeme	(Intercept)	3.013	1.736	
Trial	(Intercept)	0.040	0.201	
Number of observations: 2724				
Participants: 84				
Lexemes: 33				
Trials: 33				

University graduates produced more expected forms than high-school graduates. Longer letter strings were more likely to result in a non-expected form.

As for lexeme type, there are significant differences in expected vs. non-expected forms produced between D and N verbs, and between OA and N verbs. The difference between D and OA verbs was also significant (p-value < 0.001) according to a post-hoc simultaneous tests for general linear hypotheses with adjusted p-values.

Other variables, such as the type of task (active to passive voice vs. past to non-past tense), age, neighbourhood density of the forms, did not have a significant effect on the selection of forms. In contrast to the nouns, corpus frequency measures (e.g., Expected form fq 1) were not significant in this analysis. We will return to this point in the discussion.

4.2.4 Verbs. Expected forms only (those with a stem change vs. those without a stem change)

We replicated the analyses shown in the previous sub-section (4.2.3), this time including only *expected* inflected verb forms which constituted 75.8 % of all responses. The results of the model are presented in Table 7.

Table 7. Inflectional choices model for expected variant forms (verbs)

Fixed effects	Estimate	Std.Error	*z*-value	Pr(>|z|)
(Intercept)	31.868	5.678	5.613	<0.001
Lexeme D	-26.710	8.254	-3.236	0.001
Lexeme OA	15.449	6.794	2.274	0.023
Length	-6.202	0.638	-9.716	<0.001
Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	103.1	10.152	
	Length	2.365	1.538	-1.00
Lexeme	(Intercept)	312.3	17.671	
Trial	(Intercept)	<0.001	0.001	
Number of observations: 2098				
Participants: 84				
Lexemes: 33				
Trials: 33				

After we removed non-expected forms, unlike for the nouns (see Table 5), Lexeme type (D, OA, or N) is still a significant predictor of exactly which expected form was produced. D-type verbs were more likely to trigger those inflectional variants that require no-stem-changes (compared to N-type verbs), while OA-type verbs were more likely to trigger inflectional variants with the stem changes (compare to N-type verbs and also D-type verbs). This suggests that people prefer to produce variants without stem changes for D verbs if those are possible but apply stem changes while inflecting OA verbs when possible.

Education or corpus frequencies were not significant predictors.

4.3. Reaction times

Our third hypothesis suggested that respondents would take longer to produce answers for D and OA slots than N slots. We looked at *final reaction times* (the amount of time taken from landing on the screen to finalising the answer) and *initial reaction times* (the amount of time elapsed between landing on the screen and starting to input an answer). Between them, these two measures give a good picture of responses. Initial reaction times can show hesitance in beginning a response, which might be indicative of uncertainty; however, they cannot show whether an individual hesitates partway through the response or rethinks and repairs, which could be further indicators. Final RTs include any hesitations or delays during the response, but might also show an effect of typing speed.

4.3.1 Nouns. Final Reaction Times

For nouns, we turn first to our model for Final RT in Table 8. Here, only Lemma frequency, Length (the number of letters in the string the participant typed in each answer), and the Lexeme types (N, OA, D) significantly predicted RTs. Lemma frequency suggests that familiarity might prompt a quicker reaction; Length obviously connects with the duration needed to complete the answer. Both these effects are therefore expected. There is a significant difference in RTs between D and OA slots (according to a post-hoc simultaneous tests for general linear hypotheses with adjusted p-values) as well as between D and N slots, but not between OA and N slots. In other words, participants were slower in completing forms in defective slots in comparison to either OA or N slots. Age, gender, terms of interaction, neighbourhood density and other variables were not significant predictors of RTs.

Table 8. Final RTs for nouns

Fixed effects	Estimate	Std.Error	z-value	Pr(>|z|)
(Intercept)	9.119	0.14	65.344	<0.001
Lemma fq. (log)	-0.034	0.016	-2.104	0.041
Lexeme D	0.296	0.087	3.418	0.001
Lexeme OA	0.1	0.063	1.582	0.12
Length	0.032	0.003	9.676	<0.001
Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.171	0.413	
	Trial	<0.001	0.008	-0.33
Lexeme	(Intercept)	0.031	0.175	
Residual		0.157	0.396	
Number of observations: 3015				
Participants: 60				
Lexemes: 51				

4.3.2 Nouns. Initial Reaction Times

Initial RTs show a similar pattern. Of all the predictors, only Length (of the string the participant was planning to type), and Lexeme type significantly predicted initial RTs. For Lexeme type, again D cells are initiated more slowly than OA and N cells. The difference between OA and N cells is not significant. Perhaps surprisingly, lemma frequency of a word did not show up as a significant predictor. Participants had slower Initial RTs when they used mobile devices instead of computers; however, as these constituted only 8 out of 60 participants, this factor has not been included in the final version of the model.

Table 9. Initial RTs for nouns

Fixed effects	Estimate	Std.Error	z-value	Pr(>|z|)
(Intercept)	8.533	0.068	124.812	<0.001
Lexeme D	0.301	0.063	4.761	<0.001
Lexeme OA	0.1	0.063	1.591	0.118
Length	0.012	0.003	4.773	<0.001
Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.147	0.383	
	Trial	<0.001	0.008	-0.07
Lexeme	(Intercept)	0.031	0.177	
Residual		0.145	0.38	
Number of observations: 2999				
Participants: 60				
Lexemes: 51				

Comparisons between the three noun types for Initial RTs reveal that there is a significant difference between D and OA slots (according to a post-hoc simultaneous test for general linear hypotheses with adjusted p-values) as well as between D and N slots, but the difference between OA and N slots is not significant. In other words, participants were slower when initiating responses for the defective nouns in comparison to either OA or N nouns.

4.3.3 Verbs. Final Reaction Times

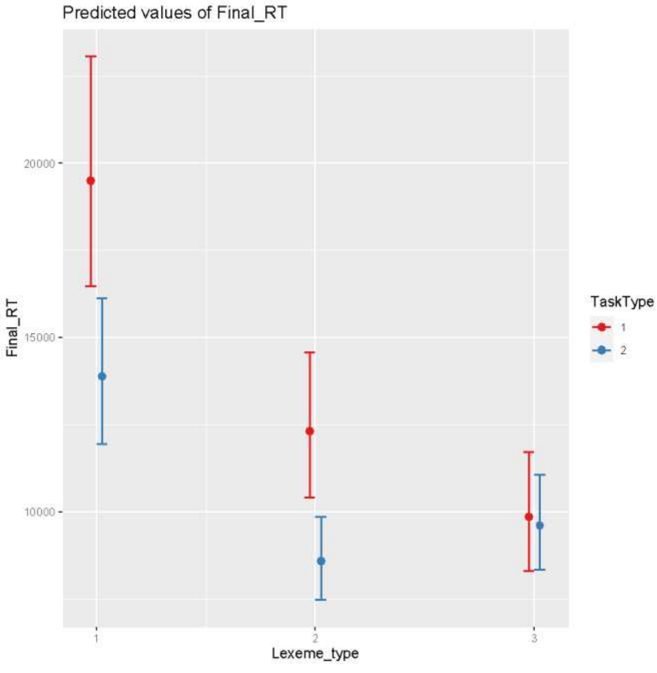
The model for final reaction time for verbs is in Table 10. As mentioned earlier, there are two different task types: active > passive and past > non-past for this group. This has been entered as a fixed effect and as an interaction with the verb type (D, OA, N), and found to be significant in the Final RT model. More effects appear in the verbal data as significant; alongside Length and Lexeme type, we also find that Age and Neighbourhood density (which can relate to issues like productivity/class size) play a role: older participants are slower to complete the task and verbs with more phonological neighbours had longer RTs than verbs with fewer phonological members. We will return to these findings in the discussion.

Initial RTs for D verbs were longer than for N verbs. Likewise, initial RTs for OA verbs were (marginally) longer than for N verbs. RTs for D verbs were longer than those for OA verbs (according to a post-hoc simultaneous test for general linear hypotheses with adjusted p-values).

Table 10. Final RTs for verbs

Fixed effects	Estimate	Std.Error	*z*-value	Pr(>|z|)
(Intercept)	8.613	0.138	62.455	<0.001
Lexeme D	0.682	0.108	6.335	<0.001
Lexeme OA	0.222	0.108	2.066	0.049
Length	0.047	0.005	8.618	<0.001
Task Type 2	-0.027	0.095	-0.280	0.782
Age	0.006	0.002	2.547	0.013
Neighborhood density	0.044	0.013	3.472	0.002
Lexeme D : Task Type 2	-0.313	0.137	-2.279	0.031
Lexeme OA : Task Type 2	-0.334	0.134	-2.499	0.019
Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.134	0.366	
	Length	<0.001	0.021	-0.36
Lexeme	(Intercept)	0.021	0.144	
Trial	(Intercept)	0.01	0.098	
Residual		0.145	0.381	
Number of observations: 2666				
Participants: 84				
Lexemes: 33				
Trial: 33				

Chart 3. Interaction between Task type (1 – tense, 2 – voice) and Verb type (1 – D, 2 – OA, 3 – N).



Task 1 (tense) causes longer RTs than Task 2 (voice) in D items and OA items, but not in N items.[15]

[15] It might seem counterintuitive that a complex operation such as active > passive was faster for respondents than past > non-past, but given that the experimental set-up focused exclusively on producing a single morphological form, it is explainable. Respondents only had to insert a form into an already created sentence, meaning that other aspects of the active > passive conversion (such as case modifications) were irrelevant, and the transformation produced a sentence identical in meaning to the prompt, whereas the past > non-past transformation involved a change in grammatical meaning.

4.3.4 Verbs. Initial Reaction Times

The model seen in Table 11 for Initial RT for verbs is similar to Final RT. One exception is that Education is also a significant predictor (university graduates were faster in their initial RT than high school graduates). Otherwise the same predictors are significant with the same direction of influence as in Final RTs. As to Lexeme type, there are significant differences in initial RT between D and OA verbs (according to a post-hoc simultaneous tests for general linear hypotheses with adjusted p-values), and between D and N verbs, but no difference between OA and N verbs

Table 11. Initial RTs for verbs

Fixed effects	Estimate	Std.Error	*z*-value	Pr(>|z|)
(Intercept)	8.359	0.139	60.096	<0.001
Lexeme D	0.654	0.121	5.402	<0.001
Lexeme OA	0.189	0.121	1.560	0.131
Length	0.021	0.005	4.139	<0.001
Task Type 2	-0.042	0.107	-0.397	0.695
Age	0.007	0.002	3.253	0.002
Neighborhood density	0.048	0.014	3.373	0.002
Tertiary education	-0.182	0.07	-2.608	0.011
Primary education	-0.047	0.148	-0.314	0.755
Lexeme D : Task Type 2	-0.345	0.154	-2.236	0.034
Lexeme OA : Task Type 2	-0.497	0.150	-3.3	0.003
Random effects				
Groups	Name	Variance	Std.Dev.	Corr
Participant	(Intercept)	0.092	0.303	
	Length	<0.001	0.016	-0.3
Lexeme	(Intercept)	0.027	0.164	
Trial	(Intercept)	0.007	0.084	
Residual		0.152	0.39	
Number of observations: 2660				
Participants: 84				
Lexemes: 33				
Trial: 33				

5. Discussion

Our null hypotheses were that D, OA and N cells would provoke indistinguishable reactions in our respondents in terms of the variety and distribution of answers provided and the reaction time measures. The results disprove the null hypotheses in part, but not entirely. We have clear evidence that our respondents treat defective cells differently from non-defective cells: they produce more variant forms for these cells, are less likely to settle on an expected form from one of two established stems, and react more slowly, both in terms of their initial thinking time (before starting to type) and their production time (from start to finish). However, we found less evidence that overabundant cells were treated differently from non-variant cells. In some instances, there were no significant differences between them, although we did find significant differences in the treatment of "expected" answers for verbs and in the initial RTs for verbs. This could have been because in some instances, any differences were better explained by other factors in our models, such as frequency. We did see a clear difference between OA and N cells in the failure to converge in the former on a single primary variant.

Overall, the lack of consistent differences between OA and N cells might in part be because the current study focuses exclusively on production; a study with a receptive component might find more differences between them. It is also, however, worth considering the overall context found in our results. In section 4.1, we remarked that, even when obvious typing errors and technical faults are excluded, triggers that should reliably prompt the same answer in 100% of instances do not always do so. Instead of having a clear division into one response – multiple responses – no responses, we see a cline from one to the other, with defective items having the greatest variety of answers and the least agreement on any one answer.

A second point, noted in section 4.1, is that sometimes the most popular answer is not an accepted form of the lexeme, or is not even a form of that lexeme. This happened six times, all with verb forms posited as defective, as can be seen in Table 12.

Table 12. Innovation and suppletion in most popular verb forms

Lexeme	Expected	Actual	No.
mít have-INF 'to have'	měta have-PASS-F.SG.NOM 'had'	vlastněna 'possess- PASS-F.SG.NOM' 'possessed'	12 / 84
klát joust-INF 'to joust'	kleje / klá joust.NP-3.SG 'jousts'	klaje joist [sic].NP-3.SG 'joists' [sic]	25 / 84
předsevzít 'undertake-INF'	předsevezme 'undertakes-[3.SG]'	předsevze 'undertooks-[3.SG] [sic]'	34 / 84
znát 'know-INF'	znána 'known-F.SG.NOM'	známá 'familiar-F.SG.NOM'	35 / 84
vědět 'know-INF'	věděno 'known-N.SG.NOM'	známo 'familiar-N.SG.NOM'	42 / 84
najít 'find-INF'	najito 'found-N.SG.NOM'	nalezeno 'uncovered-N.SG.NOM'	60 / 84

We might have expected the forms in column 2, which are either attested in dictionaries or posited based on the available stems, to be produced, but instead we get novel forms (in the case of the non-past tense) or suppletive forms borrowed from near-synonyms (in the case of passive participles). The fact that these forms yield a plurality for some D lexemes is startling, but they are found for other lexemes as well, albeit in much lower proportions – so, for example, some users reject a plural form for the word *sklo* 'glass-NOM.SG' and instead substitute forms of a count noun, such as *oken* 'windows-GEN.PL'. This spontaneous, occasional suppletion sees its logical extension in instances like the verb *najít* above, where the participle of a different verb becomes the default for expressing passive meaning.

A troubling point from the results in 4.2 is the fact that lemma and form frequency play a significant role for nouns, but do not seem to do so for verbs. There seems to be a fundamental difference between the verbal and nominal defectives in Czech, in that the latter are in general quite low-frequency, both on the lexeme and form level, at least in corpora of written Czech, whereas some of the verbs lacking passives are in fact among the most common transitive verbs in the language.[16] The frequency effects are thus likely to be in some instances an example of covariation, but as *form frequency* rather than *lemma frequency* is significant, it suggests that form frequency might come through more strongly across the board if we had a large enough corpus of spoken language.

In section 4.3 we noted occasional effects of age and education. It is not surprising that RTs are slower for older respondents; in an earlier study (Bermel, Knittl & Russell, 2018), we noted the greater experience of older respondents that results in more familiarity with a variety of forms, and at the same time a more categorical reaction against certain forms. Education also seems to prompt respondents to use more available stems of a word; this is

[16] Czechs will often defend this on semantic grounds, but the inclusion of near-synonyms that are not defective in this study shows that the semantic arguments are post-hoc.

especially true for verbs, where the use of substantially different stems is associated with more literary and formal forms than those common in the spoken language.

When we turned to the similarities between our word forms and others available in the lexicon, we found an inhibitory effect of neighborhood density on RTs, which was significant ($p = 0.002$). By contrast, Hamming Distance of One was not statistically significant in our models. The finding that the number of neighbours differing in their first phoneme is a better predictor of lexical retrieval than the number of neighbours differing in other positions has been reported in a number of other studies (e.g., Bien, Baayen, & Levelt, 2011; Vitevitch, Armbrüster, & Chu, 2004; see also Caselli et al. 2016). Previous reports of the effect of neighbourhood density on word recognition are inconsistent. Andrews (1989) found a facilitatory effect of neighbourhood density for lexical recognition, whereas Luce and Pisoni (1998) claim the effect of neighbourhood density is inhibitory due to lexical competition among phonologically similar forms. The latter claim is in line with our findings (the present study discusses word production). Therefore, the effect of neighbourhood density depends on task demands (language comprehension vs. production) and on language (e.g., Grainger & Jacobs, 1996).

6. Conclusion

A possible conclusion from our results is that *uncertainty* as a theoretical construct is simply a psychologically plausible way of describing *defectivity:* most of our findings suggest that respondents react one way to defective cells, and another way to non-defective cells, with the latter encompassing both overabundant and non-variant types.

This analysis may be in part an artefact of our production task. As distinct from a receptive task, our respondents only had to produce a single form, not suggest or evaluate all possible forms. It may be that the lexical representations deployed for production can differ from those deployed for reception and comprehension, or at least that those mental representations are accessed differently in the two modalities. In a production task, then, an OA item may not be handled any differently by most native speakers than a N item, as all they need to do is produce one form. We had isolated evidence of respondents suggesting multiple forms for OA lexemes, but this was rare.[17]

However, our previous research (see Nikolaev & Bermel, 2022) suggests that *uncertainty* is a way of looking at non-canonical material as part of a gradient category. At the beginning of this article, we said that multiple plausible stems were a frequent cause of either D or OA cells, and we sought to ascertain how respondents treated them. Our previous work focused on determining the multiplicity of forms available, which turns out to be a good proxy for the distinction between defective and non-defective cells. However, it does not account for overabundant cells, which on balance are not distinguishable from non-variant ones in this respect: even non-variant cells show some variation in form. The most salient measure for distinguishing the three phenomena instead appears to be convergence. For N cells, answers converge overwhelmingly on a single form; for OA cells, they tend to converge on one, but with substantial numbers of individuals favouring a different answer; and for D cells, there is a lack of convergence, with it being rare to find a majority favouring one form.

[17] We are grateful to an anonymous reviewer for this observation.

A model of uncertainty that incorporates these new findings would look instead like a cline of uncertainty. On one end are lexemes where few alternatives are possible, and we see only occasional deviations based on individual proclivities (a feeling that a lexeme cannot be used in a specific sort of setting – context, number, case – and must be 'covered' by a form of another lexeme). These are our classic non-variant items. On the other end are lexemes where multiple options, unfamiliarity and possibly a lack of general need yield a variety of answers and a lack of consistency in answering. These are our classic defective items. In between is a gradient of situations, where there may be multiple forms available and where convergence on a single 'best' form may be weak, or where an alternative is well-represented. An item 'in between' has in many respects the same features as a non-variant lexeme: the respondent produces it just as fast and is just as likely to produce an expected answer as a novel one.

References

Andrews, S. 1989. Frequency and neighborhood size effects on lexical access: activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15. 802–814.

Anwyl-Irvine, Alexander L. & Massonnié, Jessica & Flitton, Adam & Kirkham, Natasha & Evershed, Jo K. 2020. Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods* 52. 388–407. DOI 10.3758/S13428-019-01237-x

Baayen, R. Harald & Milin, Petar. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3. 12–28.

Baerman, Matthew. 2008. Historical observations on defectiveness: the first singular non-past. *Russian Linguistics* 32. 81–97. DOI 10.1007/S11185-007-9017-9

Baerman, Matthew. 2011. Defectiveness and homophony avoidance. *Journal of Linguistics* 47. 1-29. DOI 10.1017/S00222267-100-0002-2

Bates, Douglas & Mächler, Martin & Bolker, Ben & Walker, Steve. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67. 1–48.

Bermel, Neil. 2007. *Linguistic Authority, Language Ideology, and Metaphor: The Czech Spelling Wars.* Language Power and Social Process 17. Berlin: Mouton de Gruyter

Bermel, Neil & Knittl, Luděk. 2012. Corpus frequency and acceptability judgements: A study of morphosyntactic variants in Czech. *Corpus Linguistics and Linguistic Theory* 8. 241–275.

Bermel, Neil. 2014. Czech diglossia: dismantling or dissolution? In Árokay, Judit & Gvozdanović, Jadranka & Miyajima, Darja (eds.), *Divided Languages? Diglossia, Translation and the Rise of Modernity in Japan, China and the Slavic World,* 21–37. Dordrecht: Springer.

Bermel, Neil & Knittl, Luděk & Russell, Jean. 2015. Morphological variation and sensitivity to frequency of forms among native speakers of Czech. *Russian Linguistics* 39. 283–308.

Bermel, Neil & Knittl, Luděk & Russell, Jean. 2018. Frequency data from corpora partially explain native-speaker ratings and choices in overabundant paradigm cells. Corp*us Linguistics and Linguistic Theory* 14. 197–231.

Bien, Haidrun & Baayen, R. Harald & Levelt, Willem J. M. 2011. Frequency effects in the production of Dutch deverbal adjectives and inflected verbs. *Language and Cognitive Processes* 26. 683–715.

Booij, Geert. 2009. Morphological analysis. In Heine, Berd & Narrog, Heiko (eds.), *The Oxford handbook of linguistic analysis*, 507–529. Oxford: Oxford University Press. DOI: 10.1093/oxfordhb/9780199544004.001.0001

Caselli, Naomi K. & Caselli, Michael K. & Cohen-Goldberg, Ariel M. 2016. Inflected words in production: Evidence for a morphologically rich lexicon. *Quarterly Journal of Experimental Psychology* 69. 432–454.

Corbett, Greville G. 2005. The canonical approach in typology. In Frajzyngier, Zygmunt & udges, Adam & Rood, David S. (eds.), *Linguistic diversity and language theories*, 26–47. Amsterdam/Philadelphia: John Benjamins.

Czech National Corpus [Český národní korpus]. 2015. *SYN2015: A representative corpus of written Czech.* Prague: Faculty of Arts and Philosophy, Charles University. (http://www.korpus.cz)

csTenTen17: Corpus of the Czech Web. 2018. (https://www.sketchengine.eu/cstenten-czech-corpus)

Daland, Robert & Sims, Andrea D. & Pierrehumbert, Janet. 2007. Much ado about nothing: A social network model of Russian paradigmatic gaps. *Proceedings of the 45th annual meeting of the Association of Computational Linguistics*, 936–943. Prague: Association for Computational Linguistics.

De Smet, Hendrik & D'hoedt, Frauke & Fonteyn, Lauren & Van Goethem, Kristel. 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics* 29: 197–234. DOI 10.1515/cog-2016-0025

Gouskova, Maria & Becker, Michael. 2013. Nonce words show that Russian yer alternations are governed by the grammar. *Natural Language and Linguistic Theory* 31. 735–765. DOI 10.1007/S11049-013-9197-5

Grainger, Jonathan & Jacobs, Arthur M. 1996. Orthographic processing in visual word recognition: a multiple read-out model. *Psychological Review* 103. 518–565.

Maiden, Martin & O'Neill, Paul. 2010. On morphomic defectiveness: Evidence from the Romance languages of the Iberian peninsula. In Baerman, Matthew & Corbett, Greville G. & Brown, Dunstan (eds.), *Defective paradigms: Missing forms and what they tell us,* 103–124. Oxford: Oxford University Press (Proceedings of the British Academy). DOI: 10.5871/bacad/9780197264607.001.0001

Kováříková, Dominika. 2021. Sharing data through specialized corpus-based tools: the case of GramatiKat. *Jazykovedný časopis* 72 (2).

Kováříková, Dominika & Kovářík, Oleg. 2021. GramatiKat: A tool for research into grammatical categories and grammatical profiles. Prague: Faculty of Arts and Philosophy, Charles University. (http://www.korpus.cz/gramatikat)

Křen, Michal & Cvrček, Václav & Čapka, Tomáš & Čermáková, Anna & Hnátková, Milena & Chlumská, Lucie & Jelínek, Tomáš & Kováříková, Dominika & Petkevič, Vladimír & Procházka, Pavel & Skoumalová, Hana & Škrabal, Michal & Truneček, Petr & Vondřička, Pavel & Zasina, Adrian. 2016. SYN2015: Representative corpus of contemporary written Czech. In Calzolari, Nicoletta & Choukri, Khalid & Declerck, Thierry & Goggi, Sara & Grobelnik, Marko & Maegaard, Bente & Mariani, Joseph & Mazo, Helene & Moreno, Asuncion & Odijk, Jan & Piperidis, Stelios (eds.), *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, 2522–2528. Portorož: ELRA.

Kuznetsova, Alexandra & Brockhoff, Per, & Christensen, Rune. 2013. LmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R Package Version 2.0–33 (https://CRAN.R-project.org/package=lmerTest)

Luce, Paul A. & Pisoni, David B. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing* 19. 1–36.

Lüdecke, Daniel. 2017. sjPlot: Data visualization for statistics in social science. R package version 2.3.1 (https://CRAN.R-project.org/package=sjPlot)

Moskvin, Vasily. 2015. Недостаточные глаголы: к уточнению понятия [Defective verbs: Towards a terminological clarification]. *Russian Linguistics* 39. 209–232.

Nikolaev, Alexandre & Bermel, Neil. 2022. Explaining uncertainty and defectivity of inflectional paradigms. *Cognitive Linguistics* AOP. DOI: 10.1515/cog-2021-0041

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (https://www.R-project.org/)

Russian National Corpus [Национальный корпус русского языка]. 2003—2021. (ruscorpora.ru)

Sims, Andrea D. 2009. Why defective paradigms are, and aren't, the result of competing morphological patterns. *Proceedings from the panels of the 43rd annual meeting of the Chicago Linguistic Society.* Chicago: Chicago Linguistic Society. 267–281.

Sims, Andrea D. 2017. Slavic morphology: Recent approaches to classic problems, illustrated with Russian. *Journal of Slavic Linguistics* 25. 489–524. DOI: 10.1353/jsl.2017.0019

Suchomel, Vít. 2018. csTenTen17, a recent Czech web corpus. In Horák, Aleš & Rychlý, Pavel & Rambousek, Adam (eds.), *Proceedings of the Twelfth Workshop on Recent Advances in Slavonic Natural Language Processing,* 111–123. Brno: Tribun EU.

Thornton, Anna M. 2011. Overabundance (multiple forms realising the same cell): A non-canonical phenomenon in Italian verb morphology. In Maiden, Martin & Smith, John Charles & Goldbach, Maria & Hinzelin, Marc-Olivier (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 358–381. Oxford: Oxford University Press.

Thornton, Anna M. 2012. Reduction and maintenance of overabundance: A case study on Italian verb paradigms. *Word Structure* 5. 183–207.

Vitevitch, Michael S. & Armbrüster, Jonna, & Chu, Shinying. 2004. Sublexical and lexical representations in speech production: Effects of phonotactic probability and onset density. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30. 514–529.

Woods, Andy T. & Velasco, Carlos & Levitan, Carmel A. & Wan, Xiaoang & Spence, Charles. 2015. Conducting perception research of the internet: A tutorial review. *PeerJ* 3: e1058. https://doi.org/10.7717/peerj.1058 .