Journal of Memory and Language 107 (2019) 169–194

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Language-general and language-specific phenomena in the acquisition of inflectional noun morphology: A cross-linguistic elicited-production study of Polish, Finnish and Estonian

Sonia Granlund[a,b,c,1], Joanna Kolak[a,b,d,1], Virve Vihman[b,e], Felix Engelmann[a,b], Elena V.M. Lieven[a,b], Julian M. Pine[a,c], Anna L. Theakston[a,b], Ben Ambridge[a,c,*]

[a] ESRC International Centre for Language and Communicative Development (LuCiD), UK
[b] University of Manchester, UK
[c] University of Liverpool, UK
[d] University of Warsaw, Poland
[e] University of Tartu, Estonia

ARTICLE INFO

Keywords:
Language acquisition
Nouns
Case marking
Morphology
Cross-linguistic
Finnish
Estonian
Polish

ABSTRACT

The aim of this large-scale, preregistered, cross-linguistic study was to mediate between theories of the acquisition of inflectional morphology, which lie along a continuum from rule-based to analogy-based. Across three morphologically rich languages (Polish, Finnish and Estonian), 120 children (mean age 48.32 months, *SD* = 7.0 months) completed an experimental, elicited-production study of noun case marking. Confirmatory analyses found effects of *surface-form* (*whole-word, token*) *frequency* for Polish and Estonian, and *phonological neighbourhood density* (PND) for all three languages (using either our preregistered class-based or an exploratory form-based measure). An exploratory all-languages analysis yielded both main effects, and a predicted interaction, such that the effect of PND was greater for forms with lower surface-form frequency, which are less available for direct retrieval from memory. Cross-linguistic differences were investigated with exploratory analyses of case variance, affix syncretism and stem changes. We conclude that these findings are difficult to reconcile with accounts that posit rules or linguistic abstractions and are most naturally explained by analogy-based connectionist or exemplar accounts.

Introduction

Inflectional morphology occupies a special place in theories of both adult language processing and children's language acquisition. The considerable complexity of systems of inflectional morphology means that, in principle, this domain constitutes an ideal test case for competing theoretical accounts of the representation, processing and acquisition of language, and even of cognition more generally (Pinker, 1999).

In practice, acquisition has mostly been researched in (languages like) English with impoverished morphology. Although researchers have begun to investigate morphological acquisition in highly inflected languages (e.g., Aguado-Orea & Pine, 2015; Dąbrowska, 2004, 2008; Dąbrowska & Szczerbiński, 2006; Deen & Hyams, 2006; Janssen, 2016; Kirjavainen, Nikolaev, & Kidd, 2012; Kjærbæk, dePont Christensen, & Basbøll, 2014; Krajewski, Theakston, Lieven, & Tomasello, 2011; Laaha & Gillis, 2007; Leonard, Caselli, & Devescovi, 2002; Räsänen, Ambridge, & Pine, 2016; Saviciute, Ambridge, & Pine, 2018; Stephany & Voejkova, 2009; Stoll, Mazara, & Bickel, 2017; Xanthos et al., 2011), efforts in mediating between different theoretical accounts have been hampered by the fact that no clear pattern of findings has emerged. Furthermore, given that most previous studies investigate only a single language, it is difficult to know whether apparent cross-linguistic differences are important for theory building, or simply reflect methodological differences and/or mere sampling variability.

We seek to remedy this state of affairs by conducting a large-scale, preregistered, elicited-production study of noun case marking that is – as far as possible – identical across three languages: Polish, Finnish and Estonian. The aim is to distinguish which theoretical approaches can and cannot account for the pattern of findings observed.

* Corresponding author at: School of Psychology, University of Liverpool, Eleanor Rathbone Building, Bedford St South, Liverpool L69 7ZA, UK.
E-mail address: Ben.Ambridge@liverpool.ac.uk (B. Ambridge).

[1] Both authors contributed equally to this work.

https://doi.org/10.1016/j.jml.2019.04.004
Received 3 May 2018; Received in revised form 23 April 2019
Available online 21 May 2019

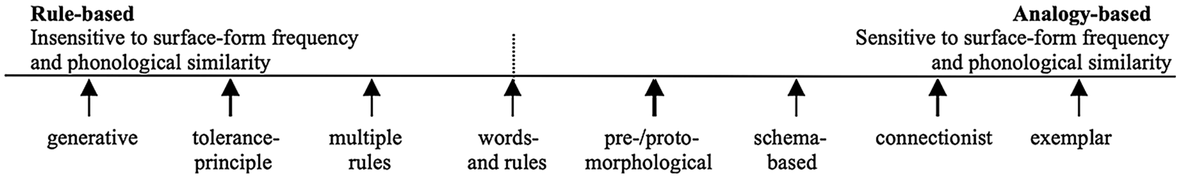


Fig. 1. A continuum of rule-based to similarity-based theoretical approaches to the acquisition of inflectional morphology.

Theories of morphological acquisition

Theoretical approaches to the acquisition of inflectional morphology (and indeed to cognitive representations more generally) can be organized along a continuum from rule-based to similarity-based (Pothos, 2005); see Fig. 1. For ease of exposition, we illustrate these approaches using the simple and well-studied example of English past-tense inflection (though as theories of morphological acquisition, all are intended to apply to verb and noun inflection generally).

At the rule-based end of the continuum lie traditional approaches based on **generative linguistics** (e.g., Deen & Hyams, 2006; Hoekstra & Hyams, 1998; Wexler, 1998), under which inflection is accomplished by formal rules that check features such as AGREEMENT and TENSE, without reference to the particular lexical item concerned (e.g., the VERB raises to the functional head TENSE, where it receives past-tense *-ed*). At the similarity end of the continuum lie **exemplar** approaches that eschew not only stored rules, but stored abstractions of any kind (see Chandler, 2010, for a review); only individual exemplars are stored (e.g., the pairs *miss → missed*, *wish → wished* and *kiss → kissed*). In production, forms are either retrieved directly from memory (in a way that is sensitive to surface-form frequency), or generated on the fly by analogy to stored exemplars, weighted by degree of phonological similarity to the target (e.g., the main analogical support for *hiss → hissed* comes from pairs such as *miss → missed* and *wish → wished*, but weaker support also comes from pairs such as *crash → crashed*, *crush → crushed* etc.).

Between these extremes lie essentially rule-based approaches that incorporate some sensitivity to statistical and phonological properties of the system, and essentially similarity-based approaches that incorporate "rules" in the sense of stored representations that abstract to some degree across individual lexical items. Yang's **tolerance principle** approach (e.g., Schuler, Yang, & Newport, 2016) posits a single rule (add *-ed*) and an inventory of stored exceptions (e.g., *sleep → slept*) that is ranked by frequency, but that is not productive; forms that are not retrieved from memory (e.g., novel or rare forms) are inflected only by means of the rule. Albright and Hayes (2003) **multiple-rules** approach posits a rule for each "island of [phonological] reliability", whether notionally regular (e.g., the rule for *miss → missed* and *wish → wished*) or irregular (e.g., *sleep → slept*, *weep → wept*). These rules are sensitive to type frequency (i.e., the number of verbs following each pattern), but not token (surface-form) frequency. The **pre-/protomorphological** approach (e.g., Bittner et al., 2003; Stephany & Voejkova, 2009) shares with Albright and Hayes (2003) model the assumption of multiple rules, but also shares with similarity-based approaches the assumption that these rules are generated by analogy, in a way that is sensitive to surface-form frequency.

The **words and rules** (dual-route) approach (e.g., Clahsen, Rothweiler, Woest, & Marcus, 1992; Pinker & Ullman, 2002) straddles the midpoint of the continuum, positing both a formal regular rule (add *-ed*) "capable of operating on any verb, regardless of its sound" (Prasada & Pinker, 1993: 2), and a similarity-based inventory of irregulars "fostering generalization by analogy" (Prasada & Pinker, 1993). **Schema-based** constructivist accounts eschew the formal regular rule at the expense of a single-mechanism system of analogy, but retain something of its flavor with the notion of a "strong schema for regular inflection" (Maslen, Theakston, Lieven, & Tomasello, 2004: 1332), or a more general goal-based schema (e.g., Bybee & Slobin, 1982: 273) such that "verbs that end in *t/d* amount to acceptable past-tense forms").

Finally, **connectionist** models of morphology (e.g., Marchman, Plunkett, & Goodman, 1997; Meunier & Marslen-Wilson, 2004; Mirković, Seidenberg, & Joanisse, 2011; Plunkett & Juola, 1999; Orsolini & Marslen-Wilson, 1997; Plunkett & Marchman, 1991,1993; Thymé, Ackerman, & Elman, 1994) – computational implementations of a broader parallel-distributed-processing approach to cognition in general (Joanisse & Seidenberg, 1999; McClelland & Patterson, 2002; Seidenberg & Gonnerman, 2000) – dispense with rules or schemas altogether, and learn individual input-output (e.g., stem-inflected form) mappings on a case-by-case basis; hidden units, however, allow for the formation of abstract representations supporting generalization to low-frequency or even novel items (e.g., see Engelmann et al., 2019, for a connectionist model of inflectional verb morphology in Polish and Finnish). Generalization can occur on the basis of any feature provided to the model; usually, in the domain of inflectional morphology, phonology and/or semantics. These patterns of hidden-unit activation are considerably more abstract than the rules or schemas posited by theories shown to the left of connectionist approaches on Pothos (2005) continuum. We have placed exemplar models (marginally) to the right because they eschew any form of stored representation whatsoever – even those as abstract as patterns of hidden unit activation – but acknowledge that this point is debatable. (The question of whether the present findings can be used to mediate between connectionist and exemplar approaches is one to which we return in the discussion.)

Differentiating the theories

Key to distinguishing these accounts of morphological acquisition are three putative effects that have been observed consistently in studies of adult processing (e.g., Baayen, Dijkstra, & Schreuder, 1997; Baayen, Kuperman, & Bertram, 2010; Baayen, McQueen, Dijkstra, & Schreuder, 2003; Baayen, Wurm, & Aycock, 2007; Baus, Costa, & Carreiras, 2008; Dąbrowska, 2008; Lõo, Järvikivi, & Baayen, 2018; Lõo, Järvikivi, Tomaschek, Tucker, & Baayen, 2018; Pham & Baayen, 2015; Vitevitch, 2002; Vitevitch & Luce, 2016; Vitevitch & Sommers, 2003), but inconsistently in studies of children's morphological production errors. The first is an effect of ***surface-form frequency*** (also known as *whole-word frequency* or *token frequency*): The greater the input frequency of the targeted inflectional form (i.e., the exact surface form that the child is attempting to produce in a given context; e.g., Polish *książki*, 'book-genitive'), the greater the speed and accuracy of production or recognition. The second is an effect of ***phonological neighbourhood density*** **(PND)**: The greater the number of "neighbours" or "friends" – nouns that are similar in both the base (nominative) form and the relevant target form (e.g., *książka → książki*; *doniczka → doniczki*; *gruszka → gruszki*) – the greater the speed and accuracy of production or recognition (though, as we will see in more detail below, phonological neighbourhoods can be defined in a number of different ways).

The third putative effect is an ***interaction of these two factors***, such that the effect of phonological neighbourhood density is greater for items with low surface-form frequency: Since low-frequency items are less likely to be successfully retrieved from memory, they must be generated by phonological analogy. These three effects, all of which we adopt as preregistered hypotheses, are consistent with theories on the similarity-based side of the continuum shown in Fig. 1. That is, exemplar models predict an effect of *surface-form frequency,* because they build in an explicit role for *exemplar storage* (i.e., rote learning of individual *ready-inflected* forms) *and retrieval*. Exemplar models predict an effect of *phonological neighbourhood density*, because they build in an explicit role for *phonological analogy to stored forms* (e.g., if *miss* → *missed* and *wish* → *wished,* then *hiss* → *hissed*). Exemplar models predict a negative interaction between the two, because phonological analogy is necessary only for lower-frequency (or novel) exemplars, which are less likely to be retrieved directly.

Connectionist models do not build in an explicit distinction between exemplar storage and phonological analogy. Nevertheless, they predict an effect of *surface-form frequency* because, holding constant other factors such as phonology, input-output (e.g., stem-inflected form) mappings that are presented to the model with high frequency are learned more readily than those presented to the model with low frequency. Connectionist models predict an effect of *phonological neighbourhood density* because the abstract representations stored in the hidden units allow the model to generalize learned input-output mappings to low-frequency or even novel forms; the more *neighbours* (or *friends*) that contribute to a particular abstract representation, the stronger – and more available for generalization – that representation will be. Connectionist models also predict *a negative interaction between surface-form frequency and phonological neighbourhood density* because high-frequency pairs (e.g., stem-output) develop their own dedicated input–output mappings, while low-frequency pairs rely more on the abstract representations stored in the hidden units. Indeed, a recent connectionist model of Polish and Finnish verb acquisition (Engelmann et al., 2019) yielded both main effects and the interaction.

The schema-based-, pre-/proto-morphological- and words-and-rules approaches predict that these effects will be observed for "irregular" parts of the system, but will be reduced, or perhaps even absent, for "regular" parts of the system (e.g., English past-tense *-ed* forms). Some accounts of this type also assume exemplar storage and phonological analogy for at least some regular forms (e.g., Alegre & Gordon, 1999; Hartshorne & Ullman, 2006). But if the regular rule, schema or route is to play any role at all (other than a merely descriptive one), it would seemingly have to predict that at least some forms are produced via this regular representation, rather than direct retrieval or phonological analogy. Furthermore, these theories have been developed largely on the basis of systems with a clear regular/irregular distinction (e.g., English past-tense, German noun-plural), and it is not clear how they would apply to systems with no clear distinction, such as those investigated here. For this reason, on our interpretation, effects of surface-form frequency, phonological neighbourhood density and their interaction are – at the very least – *less* consistent with these approaches than with connectionist and exemplar approaches.

The rule-based approaches shown to the left of the midpoint in Fig. 1 do not incorporate any mechanism that would yield an effect of surface-form frequency, or its interaction with phonological neighbourhood density. Thus, while advocates of these theories have not, to our knowledge, specifically predicted the *absence* of such effects, they could be explained – in our view – only by add-ons to the core mechanism assumed (e.g., some kind of processing advantage when applying the rule to an item to which it has been frequently applied in the past). However, all three accounts predict an effect of phonological neighbourhood density, via a role for *conjugation classes, patterns/rules* or *islands of reliability*.

It is important to distinguish here between *class-based* and *form-based* notions of phonological neighbourhood density, which we subsequently refer to as ***PND class*** and ***PND form***. Rule-based accounts assume that items such as nouns and verbs cluster into discrete conjugational or inflectional classes, which enjoy some form of representational validity. Classes are determined primarily by shared phonological properties but sometimes also by factors such as animacy and gender. Rule-based accounts predict an effect of ***PND class***, on the assumption that the strength of the rule corresponding to each class varies as a function of class-size (i.e., the number of class members). The tolerance-principle approach goes further in positing that a certain minimum class size much be reached for a rule to be stored at all. The other approaches summarized in Fig. 1 also predict an effect of PND class, to the extent that the phonological analogies and/or patterns of hidden-unit activation that they assume, approximate these classes.

The connectionist and exemplar approaches go further, in additionally positing an effect of ***PND form*** (note that our preregistration makes reference only to *PND class*; the form-based measure was devised following comments from reviewers). *PND form* starts from the notion that a purely class-based measure is potentially problematic in that classes are defined by the behaviour of a noun (or verb) across the entire morphological paradigm, rather than with reference to a particular target form. Yet children acquiring a language typically do not have knowledge of the whole paradigm, instead having access to a few selected cells, at best, and so they have no basis on which to classify words according to the classes thus defined. For example, nouns in the Polish 'Masculine1' and 'Masculine2' classes pattern identically with regard to every case context, except the locative. Thus, when the target form is not a locative form, it arguably makes little sense to treat these two classes as entirely separate. Drawing inspiration from connectionist models that map from stem to output (particularly Plunkett & Marchman, 1991, 1993; Plunkett & Juola, 1999), the measure of *PND form* (see Appendix C for details) defines "neighbourhood" on the fly, with reference to the particular target form in question. Nouns are, for a particular item, neighbours (or *friends*) if they are similar in terms of the transformation required to derive the particular target form from the nominative form.[2] For example, if the target is the Polish genitive form *mroku* ('darkness'), which has the nominative form *mrok* (i.e., *mrok* → *mroku*) then neighbours would include *krok* → *kroku* ('step'). The measure of *PND form* used in the present study also incorporates a penalty for *enemies* (e.g., Jared, McRae, & Seidenberg, 1990; Marchman, Wulfeck, & Weismer, 1999); for this example, nouns that are similar to *mrok* in nominative form, but different in genitive form, such as *smok*, 'dragon' (i.e., *mrok* → *mroku* but *smok* → *smoka*). Provided they are set up in terms of these types of mappings, both connectionist and exemplar accounts predict an effect of *PND form*. Class-based accounts would struggle to explain such an effect since they define classes with respect to the entire paradigm.

To sum up, a study that observes only an effect of *PND class* is broadly consistent with any of the theories summarized in Fig. 1. However, a study that observes an effect of *PND class and PND form*, along with an effect of surface-form frequency and its interaction with either PND measure is consistent with connectionist and exemplar accounts, but becomes increasingly difficult to explain as one approaches the rule-based end of the continuum.

What is less clear are theoretical predictions regarding interactions with age: whether effects of surface-form frequency and PND are expected to (a) decrease with age, as children build more abstract knowledge of the system (though exactly what this means is rarely specified explicitly), and so rely less on exemplar storage and phonological analogy, (b) increase with age, as stored tokens and phonological classes/forms accumulate in memory, or (c) remain relatively constant throughout development. A fourth possibility, which we adopted as our preregistered hypothesis, is (d) that the effects of surface-form frequency and PND will decrease and increase respectively

[2] Of course, in everyday speech, children do not necessarily derive case-marked noun forms from the nominative form. Purely for implementational convenience, we use the nominative form as a stand-in for those phonological properties of a noun that remain broadly consistent across the paradigm, and hence can serve as the basis for similarity-based phonological generalization. Furthermore, this implementation enjoys ecological validity for the present experimental task, in which children were presented with the nominative form of the target noun at the beginning of each trial.

with age, as children shift from reliance on rote-learned forms to phonological analogy.

Finally, not only are children's *rates* of error potentially informative for discriminating between theories, but also the *types* of errors that children produce. Some previous studies (Aguado-Orea & Pine, 2015; Dąbrowska & Szczerbiński, 2006; Dąbrowska, 2008; Leonard et al., 2002; Räsänen et al., 2016; Saviciute et al., 2018) have found that many errors involve (a) the production of a higher-frequency lexical (whole word) form in place of a lower frequency target form (e.g., Polish *książkę*, 'book-accusative' in a target context that requires *książki,* 'book-genitive'), or (b) overgeneralization of a more frequent morpheme that bears some or all of the relevant properties/features, such as gender and/or case (e.g., using the higher frequency, masculine dative marker *-owi* in place of the lower-frequency, feminine dative marker *-e*, with a feminine noun). Again, these errors are more consistent with similarity-than rule-based theories, since the former place greater emphasis on frequency-sensitive learning from the input.

Evidence from previous studies with children

Main effects of surface-form frequency and PND (specifically, PND class) are relatively well established for English, in both naturalistic (e.g., Theakston, Lieven, Pine, & Rowland, 2005; Theakston & Rowland, 2009) and experimental studies with children (e.g., Räsänen, Ambridge, & Pine, 2014; Theakston & Lieven, 2005), with many errors reflecting the substitution of higher for lower-frequency target forms (Cameron-Faulkner & Kidd, 2007; Matthews & Theakston, 2006; Räsänen et al., 2014; Theakston & Rowland, 2009) or overgeneralisation (e.g., Marchman et al., 1999; Maslen et al., 2004).

When more complex inflectional systems are considered, the picture becomes less clear. Higher surface-form frequency was found to facilitate the acquisition of noun plural marking in Danish (Kjærbæk et al., 2014), of person and number marking on verbs in Italian (Leonard et al., 2002) and Finnish (Räsänen et al., 2016), and of case marking of nouns in Lithuanian (Saviciute et al., 2018). However, Krajewski et al. (2011) found no clear effect of surface-form frequency in Polish noun morphology and Kirjavainen et al. (2012) found an effect in the opposite direction to that predicted (a negative effect) in Finnish verb morphology.

The main effect of PND (at least of PND class) is on stronger ground, having been observed in studies of noun case marking in Polish (Dąbrowska & Szczerbiński, 2006; Dąbrowska, 2004) and Lithuanian (Saviciute et al., 2018), verb inflection in Finnish (Kirjavainen et al., 2012; Räsänen et al., 2016) and verb inflection in a cross-linguistic study of Finnish and Polish (Engelmann et al., 2019). Yet differences between languages have been found in the effect of PND (class) on lexical retrieval, with dense neighbourhoods facilitating faster word finding in English, but sparse neighbourhoods leading to faster retrieval in Spanish (Vitevitch & Stamer, 2006).

However, none of the possible interaction effects set out above enjoy convincing support from previous child studies. To our knowledge, Räsänen et al. (2016) is the only study to have found interactions such that the effect of PND (class) decreases with increasing surface-form frequency and with age; moreover, the interaction of PND (class) by surface-form frequency was no longer significant when age was added to the statistical model.

The present study

In summary, despite their importance for theoretical accounts of morphological acquisition (and, indeed language acquisition more generally), the effects of surface-form frequency and PND on children's production errors remain poorly understood (particularly with regard to possible interactions). Because almost all previous studies have investigated only a single language, it is impossible to know which differences in findings are due to genuine differences between the languages under investigation versus uncontrolled differences between the studies (including, in some cases, the use of familiar versus novel lexical items and/or sampling variability). Furthermore, many of these studies have not tested the entire inflectional paradigm, but only selected subparts. This is especially the case for previous studies of Polish noun case marking (Dąbrowska & Szczerbiński, 2006; Dąbrowska, 2004, 2008; Krajewski et al., 2011). One more comprehensive study (Saviciute et al., 2018), lacked power (particularly to detect interactions), due to its small sample size.

Our goal in the present study was to remedy this situation by conducting, as far as possible, the same study of noun case marking in each of three languages – Polish, Finnish and Estonian – using a preregistered design, hypotheses and analyses, across a large, representative part of the paradigm (5–6 cases per language; see Tables A.1 and A.3 in Appendix A).

A detailed description of the Polish, Finnish and Estonian noun case-marking systems is given in Appendix A. In brief, Polish – a Slavic language – has 7 cases (nominative, accusative, genitive, dative, instrumental, locative, and vocative). In terms of PND class, Polish – according to reference grammars – has 17 declension classes (Tokarski, 1973), though with considerable syncretism across the system. Finnish and Estonian – both Finno-Ugric languages – each have 14 cases, though just three (the 'grammatical cases', nominative, genitive and partitive) make up the majority (over 64%) of surface tokens in child-directed speech corpora of both languages (described in section *Predictor variables*). Although the case-marking affixes themselves are mostly invariant across nouns, reference grammars posit as many as 51 declension classes for Finnish (e.g., *Kotimaisten kielten tutkimuskeskus*, KOTUS) and 45 for Estonian (e.g. *Eesti keele käsiraamat*, Erelt, Erelt, & Ross, 2000). In this study, we defined PND class using criteria that distinguish nouns in the targeted cases in singular form, including both noun endings and stem-changing phenomena such as vowel changes and gradation (a morpho-phonological process occurring in Finnish and Estonian, involving alternation between weak and strong syllables or phonemes; see Appendix A for more detail) and, additionally, using morphosemantic criteria in Polish (gender, animacy and countability).

The hypotheses were formulated in the preregistration (see Methods section) as follows:

1. A main effect of age: accuracy improves with age
2. A main effect of surface-form frequency[3]: the higher the surface-form frequency, the more accurate the child's performance with that particular form
3. A main effect of PND class[4]: the larger the class size, the more accurate the children's production.
4. An interaction of surface-form frequency[3] and PND class[4], such that the effect of class size will be greater for lower than for higher frequency tokens.
5. An interaction of surface-form frequency[3] and age, such that the effect of surface-form frequency will decrease with age[5].
6. An interaction of PND class and age, such that the effect of PND class[4] will increase with age[5].

We did not set out in our preregistration a plan for an analysis that collapses across all three languages and investigates possible differences between them. This was a simple oversight on our part, and should not be taken as detracting from the importance of such an analysis. Indeed,

[3] Referred to as "token frequency" in the original preregistration; changed for clarity.

[4] Referred to as "group PND" in the original preregistration, changed to "PND class" (see Appendix A) in order to contrast more straightforwardly with the subsequently-developed "PND form" measure (see Appendix C).

[5] Originally stated "will be greater for 4-year-olds than for 3 year-olds"; changed to reflect the use of age as a continuous predictor, rather than a factor.

any claims about possible differences between languages *require* such an analysis, and should not be inferred simply from patterns of significant and non-significant findings (since, frequently, "The difference between 'significant' and 'not significant' is not itself statistically significant", Gelman & Stern, 2006: 328). Unlike the individual language analyses, this analysis uses Bayesian methods only because, due to the complexity of the model, a frequentist model would not converge with anything close to maximal random-effects structure (in the sense of Barr, Levy, Scheepers, & Tily, 2013).

Finally, we also investigated, in an exploratory fashion, a number of lesser-studied, fine-grained morphological properties that vary considerably across languages (e.g., Dressler, 2003), and that are expected to affect ease of acquisition. These analyses are entirely exploratory in nature, and do not constitute tests of particular theories. Rather, they are designed to uncover potential patterns that, if subsequently confirmed, will need to be explained by future theories.

Method

Preregistration

The study was preregistered on the website of the Open Science Framework. The project site https://osf.io/bmncq/ hosts the corpus files, experimental stimuli, raw data, analysis scripts and (via https://osf.io/bmncq/registrations) the registration form (direct link: https://osf.io/3ws7a/register/565fb3678c5e4a66b5582f67).

Participants

Each of the three languages was tested on location, in nurseries in countries where the language is spoken by the majority. All tested children were reported by their parents and teachers to be typically developing, monolingual speakers of their respective languages. The age range of 3;0–5;0 was chosen on the basis of previous similar studies conducted by our group (e.g., Räsänen et al., 2016; Saviciute et al., 2018), such that children would be sufficiently old to be able to complete the task, and sufficiently young to be expected to produce errors, and drawn from a sufficiently wide age range to allow for investigation of possible developmental effects.

In line with our preregistered sampling plan ("At least 40 and a maximum of 50 participants will be tested in each language"), the initial sample comprised 40 Polish, 49 Finnish and 43 Estonian children. However, as pointed out by a reviewer, having different sample sizes and age ranges for each language makes cross-linguistic comparisons difficult. Although the inclusion of age as a continuous predictor in our statistical models addresses this problem to a certain extent, we decided to err on the side of caution and match our datasets on both factors. We therefore removed the data from the 9 oldest Finnish children and the 3 oldest Estonian children, resulting in a dataset balanced for both sample size ($N = 40$ per language, conforming to the minimum set out in the preregistration) and age: We consider the resulting span of just over two months between the mean ages of the youngest (Polish) and oldest (Finnish) groups to be acceptable.[6]

Polish children (16 female; mean age: 47.0 m, range: 31.9–58.0 m, SD: 8.2 m) were tested at a nursery located in Olsztyn, Northeast Poland. Finnish children (17 female; mean age: 49.34 m, range 36.2–59.8 m, SD: 6.6 m) were recruited from four nurseries in Pori, Southwest Finland. Estonian children (22 female; mean age: 48.7 m, range: 38.0–60.0 m, SD: 6.0 m) were tested in five nurseries in Tartu, Southeast Estonia.

[6] Although a common approach would be to compare the ages of the three language groups statistically (using a one-way ANOVA), this is statistically inappropriate, for reasons outlined in Sassenhagen and Alday (2016).

Predictor variables

In all three languages, surface-form frequency and PND counts (both PND class and PND form) were derived from child-directed speech corpora. In Polish, the child-directed speech frequency list from Haman et al. (2011) was used. The list is based on speech directed to children between 0;10 and 7;0 by adults and older children, and incorporates data from seven corpora with a total of over 794,000 word tokens. In Finnish, The Kirjavainen-Max Planck corpus (Kirjavainen, Kidd, & Lieven, 2017) was used. It contains approximately 684,000 word tokens from recordings of a mother, father and adult relatives interacting three hours per week with a child, Piia, between the ages of 1;7 and 4;0 (total 278 files). For Estonian, the frequency and PND counts were derived from the Argus (Argus, 1998), Vija (Vija, 2007), and Zupping (Zupping, 2016) corpora available on CHILDES, including 129 recordings with speech directed to children between the ages of 1;3 and 4;2 by both mothers and fathers. This included over 171,000 tokens.

Ideally, it would have been preferable to use corpora that were more comparable across languages (rather than a single-child corpus for Finnish and multiple-child corpora for Polish and Estonian; and a smaller corpus for Estonian than for Finnish and Polish). Unfortunately, no such corpora were available. In order to verify that these corpus differences were unlikely to have affected the pattern of results observed, we plotted – for each language – Zipf curves of all surface noun forms present in each corpus (not just those used as targets in the present study). These plots (see Fig. B1 in Appendix B) confirm that the distribution of case-marked nouns was similar for each language corpus.

Surface-form frequency counts were obtained automatically from the corpora (using a custom-written R script). The procedures for calculating PND class and PND form are outlined in Appendix A and Appendix C respectively. Even though adjectives share inflectional endings with nouns in Finnish and Estonian, we considered only nouns for class size and continuous-PND calculations, in order to keep the measures comparable with Polish, where adjectival and nominal inflections differ. Only singular noun forms were considered, as plurals follow a different (somewhat simpler) inflectional pattern in each language. Although our preregistration made reference only to PND class, the pattern of results obtained here does not hinge crucially on the use of one or other particular operationalization of PND.

Design and materials

The study employed an elicited-production method and within-subjects design. Due to the large number of cases in each of the languages, it was not possible for us to test children on every case. Instead, we selected the most frequently occurring singular cases, based on the child-directed speech corpora in each language. Neither was it possible to include items from every noun class (i.e., PND class). Instead, we selected three nouns from each of 10 (Polish), 9 (Finnish) or 8 (Estonian) classes (see Tables A.2, A.4 and A.5 in Appendix A for the selected classes (indicated with a tick) and Tables D.1–D.3 in Appendix D for the selected nouns).

Nouns were chosen to span a range from high to low frequency in child-directed speech (summing across all surface forms of each noun), in order to ensure that the individual elicited target forms varied continuously with regard to the predictor of surface-form frequency. Nouns were chosen to be easy to depict and known to children in the participants' age range. For detailed criteria on selecting the stimuli, see Appendix D. An additional three nouns were chosen for practice trials in each language; the practice items were highly frequent nouns from the largest class in each of the corpora.

Participants were tested on 5 (Polish) or 6 (Finnish and Estonian) different cases in the singular (see Tables 1 and 2), which resulted in 150 Polish, 162 Finnish and 144 Estonian stimuli. For Polish, one case (vocative) was not included because it is not used in colloquial speech.

Table 1
Polish elicitation contexts for each case. Note that the same contexts were used with both the fox and hare characters.

Case	Context	Translation
Genitive	Lis macha do X.	'The fox waves at X'
Dative	Lis daje prezent X.	'The fox gives a present to X.'
Accusative	Lis patrzy na X.	'The fox looks at X.'
Instrumental	Lis bawi się X.	'The fox plays [with] X.'
Locative	Lis myśli o X.	'The fox thinks about X.'

Table 2
Finnish and Estonian elicitation contexts. Note that the same contexts were used with both the fox and hare characters.

Case	Context (Finnish)	Context (Estonian)	Translation
Genitive	Kettu ottaa X.	Rebane võtab X.	'The fox takes X.'
Partitive	Kettu ei näe X.	Rebane ei näe X.	'The fox does not see X.'
Inessive	Kettu asuu X.	Rebane elab X.	'The fox lives [in] X.'
Elative	Kettu tulee ulos X.	Rebane astub välja X.	'The fox leaves [the inside of] X.'
Illative	Kettu menee sisään X.	–	'The fox goes into [the inside of] X.'
Comitative	–	Rebane mängib X.	'The fox plays [with] X.'
Allative	Kettu vilkuttaa X.	Rebane lehvitab X.	'The fox waves [at] X.'

For Finnish and Estonian, only the most frequently used cases were chosen. Because of parallel use of case-marked and postpositional forms, adessive case was excluded, despite its frequency in the child-directed speech. Additionally, illative case was excluded in Estonian because of the frequent use of two parallel forms. In lieu of illative case, in the Estonian study we included comitative case, used frequently in colloquial speech in Estonian (7.2% in our corpus data), but not in Finnish.

Each case was represented pictorially using a unique elicitation context involving a fox or hare character interacting with each target noun object (see Tables 1 and 2). Elicitation contexts were chosen for each tested case from the most frequent types of usage in the child-directed speech corpora in each language, while also considering how easily depictable the contexts were. Each character appeared in half of the trials seen by each participant. For example stimuli, see Fig. D.1 in Appendix D.

A reviewer noted that, as a result of the need to hold nouns constant across cases and scenarios, some of the stimuli, and hence associated sentences, are rather semantically implausible (e.g., a hare coming out of a book). We did consider obtaining, and partialing out statistically, semantic plausibility ratings, but decided that these judgments would almost certainly exhibit a problematic degree of collinearity with surface-form frequency. That is, it was relatively easy to come up with plausible scenarios for case-marked forms with high surface-form frequency (e.g., accusative and partitive forms) but difficult to come up with plausible scenarios for case-marked forms with lower surface-form frequency (e.g., dative, locative and instrumental forms).

The stimuli in all contexts were hand-drawn, scanned and digitally coloured. The animal characters were drawn separately from the noun objects – each noun picture was then combined with the fox and hare elicitation context pictures, to ensure that all the pictures for each context were as similar as possible. The size of the object pictures was matched, so that the various objects used in a particular context were of similar salience.

To prevent fatigue effects, each participant was tested on all nouns, but each in only three of the five (Polish) or six (Finnish and Estonian) case contexts. This resulted in 90 (Polish), 81 (Finnish) and 72 (Estonian) trials per participant. We ensured that each child was tested on all cases within each (descriptive) noun class. The stimuli in each list were pseudo-randomised between and across participants; the same noun and the same case never occurred consecutively. Each participant's master list was divided in two for use in two separate sessions.

Procedure

Each child was tested individually in a quiet setting at their nursery. The full set of trials for each participant was presented in two experimental sessions, lasting approximately 15 min each. Each experimental session contained half of the test stimuli for each child (45 trials in Polish, 40/41 in Finnish and 36 in Estonian). There was a break between the experimental sessions; the second session was conducted either a few hours after the first session or on a subsequent visit during the next few days.

The child was seated in front of a laptop computer (13-in. screen). All the pictures were presented through Processing 3.0.1, a Java-based program (www.processing.org). The experimenter presented each picture by pressing the forward button on the keyboard. Audio recordings of the sessions were made using Audacity (www.audacityteam.org).

In each experimental trial, the participant was presented with a picture of the object on the screen and told the name of the object in nominative form. S/he was then shown the stimulus picture (i.e., the object in a case elicitation context), and the experimenter produced the beginning of the context sentence, for example:

1. Experimenter: 'This is an apple-NOM.' [screen shows an apple]
2. Experimenter: 'The fox looks at…' [screen shows picture of the fox looking at an apple]
3. Participant: 'apple-ACC'

To ensure that children were familiar with all the elicitation contexts, and that any errors could not be attributed to misinterpretation of the case contexts, the two experimental sessions were preceded by a short training session lasting between 5 and 10 min. In the first training session, the child was told that s/he would be telling a story together with the experimenter; two characters would be going on an adventure and interacting with different objects in sometimes-surprising ways. First, the fox and hare were introduced to the child. Then, the participant was trained on all case elicitation contexts with two training nouns, one presented with the fox and one with the hare. Therefore 10 (Polish) or 12 (Finnish and Estonian) training trials preceded the test trials. For the first training noun, the experimenter completed all the sentences correctly using the target noun, and the participant repeated after the experimenter. For the second training noun, the participant completed the sentences independently, and the experimenter corrected any mistakes.

In the second practice session, at the beginning of the second experimental session, the participants were trained on one noun in all contexts using one character only (i.e., 6 training trials in Finnish and Estonian, and 5 training trials in Polish) to ensure that they remembered the task and contexts correctly. Again, the experimenter corrected the participant if necessary.

In the experimental sessions, the procedure was the same as in the practice sessions, but the experimenter did not correct the participant, and provided only positive feedback. If the participant produced an incorrect noun (regardless of the case form produced), the experimenter reminded him/her of the name of the stimulus (i.e. 'Do you remember what this is? This is a X.') and provided the elicitation context again. If the child still did not provide the correct noun in his/her response, the experimenter articulated the first sound in the noun to encourage the child to respond with the correct noun (i.e., 'The fox looks at the a-'). If the child still produced an incorrect noun, the experimenter moved on to the next trial.

To keep participants motivated throughout the experiment, at selected intervals during each session, the participants received colourful tokens, which, at the end of each session, were exchanged for a sticker.

Table 3
Percentage of each type of response in the three languages. Scorable responses (1 = correct, 0 = incorrect) and five types of unscorables.

%	1	0	No response	Diminutive	Prepositional/postpositional phrase	Wrong noun	Partitive for genitive
Polish	70.4	15.8	1.1	0.3	12.2	0.1	–
Finnish	78.3	15.7	2.7	–	2.9	0.4	–
Estonian	75.4	8.5	0.8	0.2	11.7	0.3	3.1

Transcription and coding

During each test session, the experimenter transcribed the participant's response to each trial. Any unclear responses were marked as such and, later, the audio recording was consulted for clarification. A total of 3466 Polish, 3025 Finnish and 2880 Estonian responses were produced by the participants.

All responses were coded as unscorable (NA), correct (1) or incorrect (0). Responses were considered unscorable if the participant produced no response, an unintelligible response, a diminutive or augmentative form, a prepositional or postpositional phrase, a stem on its own without an inflection (in Polish only), a noun other than the target, or a specific partitive for genitive substitution (in Estonian only). Some previous studies (e.g., Kempe, Brooks, Mironova, Pershukova, & Fedorova, 2007; Laalo, 2001; Savickienė, Kempe, & Brooks, 2009) have found diminutivisation to be a common strategy used by children to fill gaps in their morphological knowledge, since these forms tend to be highly regular. However, diminutive forms were produced very rarely in the present study (< 0.4% for each language), presumably because children were supplied with the target noun in non-diminutive nominative form on each trial. We also marked as unscorable any responses which were well-formed in the language using means other than the targeted case; these included pre- and postpositional phrases and responses in which the child produced a partitive form in place of a genitive target. We comment further on these in the *Descriptive error analysis* section, below. Table 3 presents the percentages of scorable responses and the various categories of unscorable responses in each language.

In all languages, scorable responses were classified as either correct inflection (1; the participant produced the correct target for the context) or incorrect inflection (0; the participant produced an incorrect target form for the context, regardless of whether the error occurred in the stem or the affix). Consonant gradation errors were not considered errors in Finnish because consonant gradation is a general phonological process in Finnish, meaning that errors of consonant gradation are not errors of case marking per se. In Estonian, on the other hand, gradation[7] is integral to case formation, and in some cases constitutes the only cue to case; it is not only phonologically conditioned. Hence, errors in consonant gradation were included in assessing response accuracy in Estonian.

In order to calculate reliabilities, 15–17% of the responses were transcribed independently by another native speaker of each of the languages. Agreements were 99% in Polish, 92% in Finnish and 96% in Estonian. For all the analyses in the Results section, the unscorable responses were disregarded – only the correct (1) and incorrect (0) responses were analysed.

Results

The effect of input-based predictors

First, we conducted the preregistered confirmatory analysis using linear mixed-effects models to investigate whether children's accuracy in noun case marking in each of the three languages could be predicted by the children's age, surface-form frequency and PND class.

Statistical methodology

This analysis is confirmatory in the sense that we have set out our hypotheses and the method of analysis in a preregistration on the website of the Open Science Framework (see start of Methods section) before the beginning of data collection. Linear mixed-effects regression models were run according to the preregistration, deriving beta coefficients and *p*-values. Mainly because of convergence issues with the frequentist models we also computed Bayesian posterior distributions for the same models. We explain the methods of both approaches and the differences between them below.

In order to account for item and participant variance, the frequentist analysis was conducted using linear mixed effects regression models (Baayen, Davidson, & Bates, 2008) with the lme4 package (Bates, Machler, Bolker, & Walker, 2015) in R (R Core Team, 2016). A generalized mixed effects model with a binomial link function was fitted with response accuracy as a binary dependent variable (0, 1) and fixed effects of (i) age (in months), (ii) target surface-form frequency and (iii) PND class. Surface-form frequency and class size were log-transformed, scaled and centred, and age was scaled and centred to 48 months in all three languages. All interactions were included in the model, with the exception of the three-way interaction of age × surface-form frequency × class size, for which our design lacked power according to prior power estimates. Random intercepts for participant and noun and by-participant random slopes were specified. Maximal models were run initially and, in cases of non-convergence, the random structure was simplified using the procedure outlined in Barr et al. (2013). P-values were obtained using ANOVA model comparison.

Our method for follow-up analyses on significant interactions slightly differed from our specified preregistered analyses: although our preregistration determined that we would subset any significant interaction data and rerun the linear mixed-effects models, we used interaction plots instead to interpret the effects. Because all our factors were continuous, subgrouping the factors, for example to "low" or "high" surface-form frequency, did not seem as reliable as viewing them as a single continuous factor.

In addition to the frequentist analysis, we fitted Bayesian generalized linear mixed-effects models with a binomial link function using the *brms* package (Bürkner, 2018) in *R*, which runs *RStan* (Stan Development Team, 2015) in the background. The use of Bayesian model fitting has several advantages (see Nicenboim & Vasishth, 2016 for an overview). A major reason for using Bayesian models is that a maximal random effects structure (Barr et al., 2013) can be fitted without convergence problems. Another advantage of Bayesian inference is that it allows the computation of "credible intervals", which have a more straightforward interpretation than the often-misunderstood confidence intervals (see Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016); credible intervals provide the range within which the true effect lies with a certain probability given the data. Despite their obvious advantages, there is still hesitation in the field about fully adopting and supporting Bayesian methods. We therefore report both the frequentist and Bayesian results, and base our conclusions on both of them.

In the Bayesian analyses, we always fitted maximal models as justified by the design (random intercepts for participant and noun and by-participant random slopes). We used weakly informative priors for fixed

[7] Gradation occurs on the level of the syllable in Estonian, affecting both consonants and vowels. However, in our stimuli, we only selected words with consonant gradation.

and random effects (i.e., we did not impose any prior information on the estimates).[8] We report the mean estimate, the lower and upper limits of the 95% credible interval, and the probability of the effect being smaller than (for negative estimates) or greater than (for positive estimates) zero (P_{MCMC}). Note that the probability P here in the Bayesian sense can be interpreted literally as the probability of the true effect being different from zero in the direction of the estimated mean, given the data, and is, thus, fundamentally different from the *p-value* in the sense of null hypothesis significance testing (NHST), which is the probability of an effect of at least the observed magnitude, given that the null hypothesis is true.

As there is no binary decision threshold for significance in the Bayesian approach, we interpret the Bayesian results as follows: If the probability $P < 0$ or $P > 0$ is close to 0.5, we conclude there is no evidence for an effect; if the credible interval contains zero but P is relatively high, we say that there is weak evidence, and when P > = 0.95 or the credible interval does not span zero, we interpret this as strong evidence for an effect given the data.

It is important to note that using both frequentist and Bayesian approaches does not amount to what is known as *p-hacking*. The method is by definition more conservative, in that (a) we do not assume strong evidence for an effect unless this conclusion is suggested by *both* methods of analysis and (b) the Bayesian method always uses full random effects structure, which is maximally conservative (Barr et al., 2013).

Polish

In Polish, the final converged frequentist model, shown in Table 4, included random intercepts for participants and items (noun lemmas), as well as by-participant random slopes for surface-form frequency. The frequentist analysis showed significant main effects of PND class, surface-form frequency and age, as well as an interaction between PND class and surface-form frequency. The Bayesian analysis, which used a maximal random structure, supported most of these results, though not with strong evidence in all cases: While the Bayesian model showed strong evidence for the main effects of surface-form frequency and age, the evidence for an effect of class size and its interaction with surface-form frequency was weaker. In summary, both the frequentist and Bayesian results suggest that accuracy improved, as predicted, with age, increasing surface-form frequency and increasing class size. A negative interaction between class size and surface-form frequency indicated that the effect of class size was more pronounced for low-frequency forms (see Fig. 2).

Finnish

In Finnish, the final converged frequentist model shown in Table 5 included random intercepts for participants and items, as well as by-participant random slopes for age and for an interaction between surface-form frequency and age. The frequentist analysis showed significant main effects of PND class and age. This was supported with strong evidence for an effect of age in the Bayesian analysis and some evidence for an effect of class size. Additionally, the Bayesian model showed very weak evidence for an effect of surface-form frequency. There is no evidence for an interaction between class size and surface-form frequency (also see Fig. 3).

Estonian

The frequentist model for Estonian (see Table 6) only converged when all random slopes were excluded, with only random intercepts for participants and items retained. Both frequentist and Bayesian analyses revealed clear main effects of PND class, surface-form frequency and age. In addition, the Bayesian model showed some, but not strong, evidence for a class size by surface-form frequency interaction in the predicted direction (see Fig. 4).

Summary

The preregistered frequentist analyses demonstrated varying amounts of evidence for effects of surface-form frequency and class size across the three languages, with class size significant in all languages, and surface-form frequency significant in Polish and Estonian. The interaction of the two factors was significant only for Polish.

Figs. 5–7 show Bayesian means and credible intervals for the effects of interest across the three languages. The effect of class size was positive in all languages, but supported by strong evidence only for Estonian. Evidence for the effect of surface-form frequency was strong for Polish and Estonian, but weak for Finnish. There was some evidence for a negative interaction between class size and surface-form frequency for Polish and Estonian. For Finnish, however, with a mean estimate close to zero and a rather narrow 95% credible interval, the result can be interpreted as positive evidence of the *absence* of an interaction.

In order to determine whether input-related factors affected children's performance in the same way in the three languages (even if they did not always cross the threshold into significance), we conducted an additional exploratory analysis by pooling the data from all three languages. As well as probing for possible differences between languages, this pooled analysis also addresses the potential objection that, although our sample size of $N = 120$ is relatively large overall, it is not particularly large for any individual language ($N = 40$), particularly given the relatively low rates of error observed (which can give rise to ceiling effects).

Pooled exploratory analysis

Because of its exploratory nature, and because the equivalent frequentist model would not converge with close to maximal random effects structure, the pooled analysis was done using the Bayesian method only. We used the same model structure as in the confirmatory analyses above, except that we added a fixed effect of language and its interactions with surface-form frequency and PND class, and excluded age. Because the language variable had three levels, it was coded with Helmert contrasts, such that Finnish was compared to Estonian, and Polish was compared to the mean of both the other languages. This is preferable over a treatment contrast that uses one language as the reference level to the others, in the interest of avoiding multiple comparisons. Comparing Polish to the mean of Finnish and Estonian is justified by the fact that Finnish and Estonian are linguistically related to each other, but not to Polish. Age was excluded because four-way interactions are difficult to interpret, and age effects were observed in each of the individual by-language models.

Results are presented in Table 7; Fig. 8 provides a more accessible visual representation, showing mean estimates and 95% credible intervals for all fixed effects. We consider an effect to be present with strong evidence when the credible interval does not span zero.

Helmert contrasts on the main effect of language revealed some evidence for poorer accuracy for Finnish than Estonian and strong evidence for poorer accuracy in Polish than the other two languages. However, these contrasts should be interpreted with caution, given that the credible intervals shown in Fig. 9 suggest that the only meaningful difference is higher accuracy for Estonian than Polish and Finnish.

The pooled analysis shows strong evidence for main effects of PND class, surface-form frequency and age across the languages. Positive coefficients imply that, as predicted, across languages, accuracy improved with larger class sizes, higher surface-form frequency and age. As described in the previous sections, Bayesian analyses found some evidence for a main effect of PND class for all three languages, but strong evidence was found only for Estonian. Helmert contrasts in the

[8] The priors for the intercept and slope were a Student *t*-distribution with 2 degrees of freedom, centred on 0. For the random-effects correlation matrix, a so-called LKJ prior was used (see the documentation in Stan Development Team, 2018; for a tutorial, see Sorensen, Hohenstein, & Vasishth, 2016).

Table 4
Results of frequentist and Bayesian analyses for overall correct responses in Polish. For the Bayesian analysis, the estimated mean is reported as well as the lower and upper limits of the 95% credible interval and the probability *P* of the effect being < 0 (for negative coefficients) or > 0 (for positive coefficients), given the data. Values in boldface indicate significance at p < 0.05 (frequentist) and strong evidence (Bayesian P ≥ 0.95).

	Frequentist				Bayesian			
Comparison	Est.	SE	*z*	*p*	Mean	Lower	Upper	*P*
Intercept	1.90	0.17	–	–	1.94	1.54	2.35	1
PND class	**0.28**	**0.12**	**3.19**	**.003**	1.62	−0.56	3.92	.93
Surface-form frequency	**0.26**	**0.08**	**3.07**	**.003**	**0.46**	**0.16**	**0.75**	**1**
Age	**0.62**	**0.13**	**4.81**	**< .001**	**0.61**	**0.31**	**0.90**	**1**
PND class:Surface-form frequency	**−0.23**	**0.06**	**−3.52**	**< .001**	−1.1	−2.96	0.78	.88
PND class:Age	0.02	0.05	0.38	.707	0.04	−0.98	1.08	.54
Surface-form frequency:Age	0.03	0.07	0.55	.587	−0.01	−0.27	0.25	.54

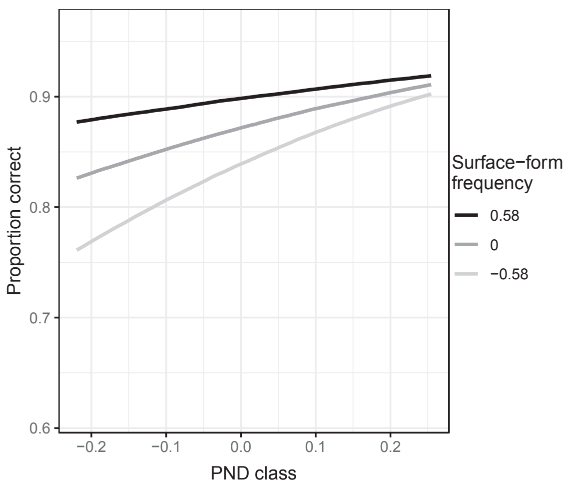


Fig. 2. Effect of log PND class in Polish as conditioned on different values of log surface-form frequency (mean ± 1SD). Note: The interaction plots show the effect on performance (y-axis) when predictor 1 (x-axis) is conditioned on predictor 2 (grayscale legend). Predictor 2 is split into three levels: the mean and 1 SD above and below the mean. All other factors are kept constant at their mean. Axes and legend show predictor values after scaling and centring.

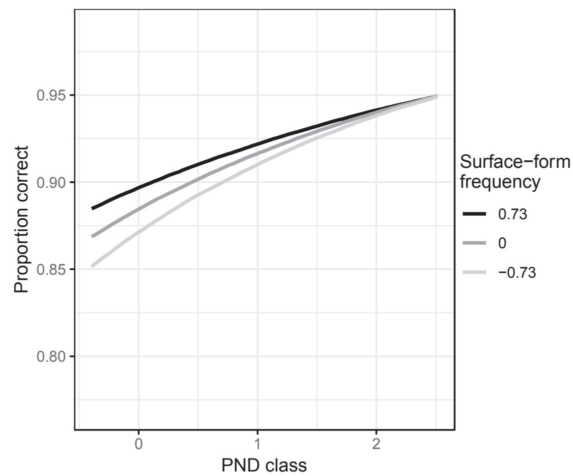


Fig. 3. Effect of log PND class in Finnish as conditioned on different values of surface-form frequency (mean ± 1SD).

Table 5
Results of frequentist and Bayesian analyses for overall correct responses in Finnish. For the Bayesian analysis, the estimated mean is reported as well as the lower and upper limits of the 95% credible interval and the probability *P* of the effect being < 0 (for negative coefficients) or > 0 (for positive coefficients), given the data. Values in boldface indicate significance at p < 0.05 (frequentist) and strong evidence (Bayesian P ≥ 0.95).

	Frequentist				Bayesian			
Comparison	Est.	SE	z	*p*	Mean	Lower	Upper	*P*
Intercept	1.85	0.23	–	–	1.93	1.42	2.47	1
PND class	**0.37**	**0.18**	**2.07**	**.047**	0.38	−0.11	0.87	.94
Surface-form frequency	0.07	0.09	0.78	.442	0.16	−0.12	0.44	.87
Age	**0.52**	**0.18**	**2.94**	**.029**	**0.53**	**0.12**	**0.90**	**.99**
PND class:Surface-form frequency	−0.04	0.07	−0.59	.510	−0.07	−0.32	0.20	.72
PND class:Age	−0.05	0.06	−0.73	.471	−0.12	−0.33	0.11	.86
Surface-form frequency:Age	0.04	0.09	0.52	.608	0.05	−0.20	0.31	.64

pooled analysis suggest strong evidence for the main effect of class size being stronger in Estonian than in Finnish (see Fig. 10), but that Polish did not differ from these languages.

In the analyses described in the previous sections, the main effect of surface-form frequency was not found in Finnish. In the pooled analysis, Helmert contrasts revealed strong evidence for the main effect of surface-form frequency being smaller in Finnish than in Estonian, and largest overall in Polish (see Fig. 11), confirming the pattern suggested by Fig. 6.

Additionally, the pooled analysis showed an interaction between surface-form frequency and PND class that lies on the threshold between some evidence and strong evidence. The negative interaction coefficient suggests that, as predicted, the effect of class size increased with decreasing surface-form frequency across languages (see Fig. 12). Helmert contrasts showed weak evidence for the interaction being stronger in Polish than in Finnish and Estonian. As discussed above, the credible interval of the PND class by surface-form frequency interaction was centred on zero in the Finnish single-language analysis. We therefore interpret this as evidence of absence of an effect in Finnish, even though the pooled model does not reveal strong evidence for differences in the interaction between languages (the 95% credible intervals in Polish and Estonian are too large to detect a difference).

Table 6
Results of frequentist and Bayesian analyses for overall correct responses in Estonian. For the Bayesian analysis, the estimated mean is reported as well as the lower and upper limits of the 95% credible interval and the probability *P* of the effect being < 0 (for negative coefficients) or > 0 (for positive coefficients), given the data. Values in boldface indicate significance at $p < 0.05$ (frequentist) and strong evidence (Bayesian $P \geq 0.95$).

	Frequentist				Bayesian			
Comparison	Est.	SE	z	*p*	Mean	Lower	Upper	*P*
Intercept	3.47	0.37	–	–	3.67	2.81	4.59	1
PND class	**0.83**	**0.30**	**2.79**	**.010**	**3.14**	**0.08**	**6.40**	**.98**
Surface-form frequency	**0.38**	**0.12**	**3.08**	**.001**	**0.41**	**0.02**	**0.84**	**.98**
Age	**0.92**	**0.24**	**3.84**	**< .001**	**1.08**	**0.39**	**1.92**	**1**
PND class:Surface-form frequency	0.10	0.11	0.92	.368	−0.91	−2.33	0.47	.90
PND class:Age	0.04	0.11	0.41	.693	0.17	−1.20	1.65	.59
Surface-form frequency:Age	−0.02	0.12	−0.17	.871	−0.09	−0.50	0.34	.68

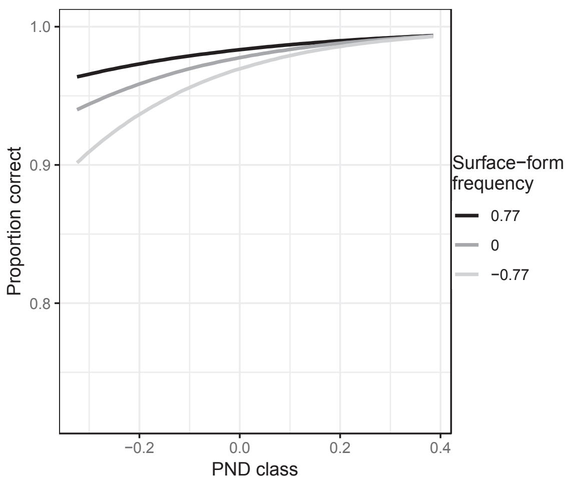


Fig. 4. Effect of log PND class in Estonian as conditioned on different values of surface-form frequency (mean ± 1SD).

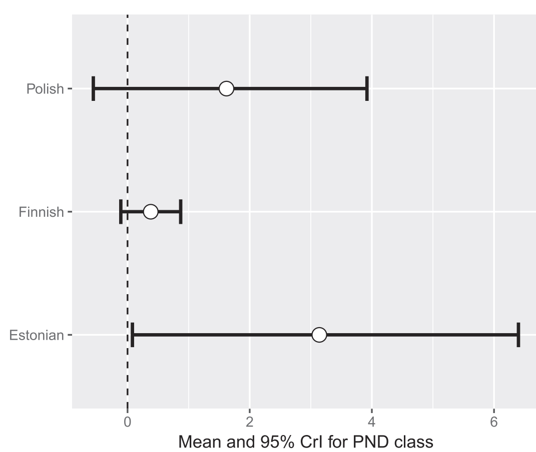


Fig. 5. Bayesian mean estimates and 95% credible intervals for the effect of PND class in all three languages.

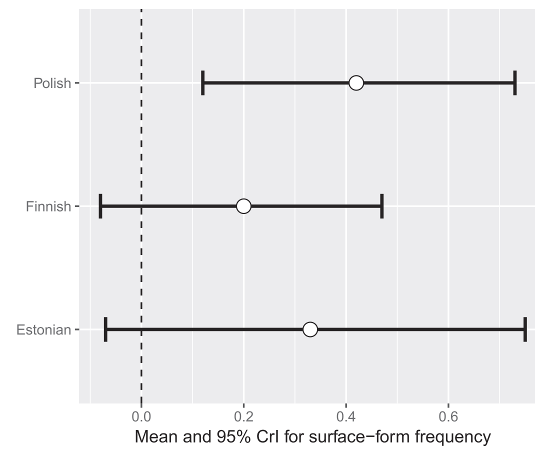


Fig. 6. Bayesian mean estimates and 95% credible intervals for the effect of surface-form frequency in all three languages.

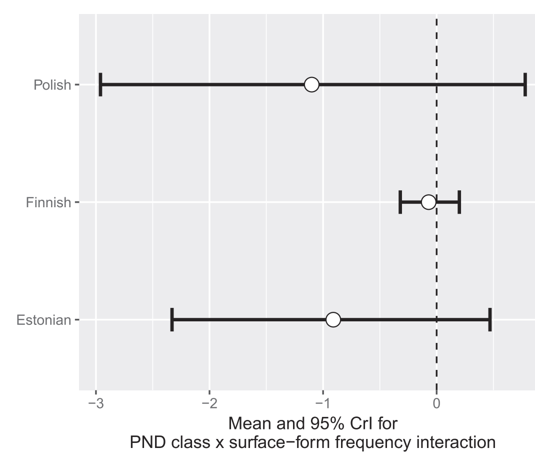


Fig. 7. Bayesian mean estimates and 95% credible intervals for the interaction between log PND class and log surface-form frequency in all three languages.

Summary

The confirmatory statistical analyses demonstrate pervasive support for the predicted positive effect of class size for all languages, and positive effect of surface-form frequency for Polish and Estonian, but not for Finnish (see Table 8 for a summary). The predicted negative interaction between surface-form form frequency and class size, reflecting an effect of phonological analogy only for lower frequency forms, was observed only for Polish. The exploratory pooled analysis found that, although some of these cross-linguistic differences may be meaningful, the predicted main effects and negative interaction of surface-form frequency and class-size hold when collapsing across languages in order to achieve a well-powered analysis. Finally, although main effects of age were pervasive, we found no evidence in any analysis for any interaction of age by surface-form frequency or class size.

Analysis using continuous, form-based PND measure (PND form)

Finally, we also conducted an exploratory analysis using a more sophisticated, continuous measure of PND, *PND form*, which is based on the individual noun form instead of the declension class membership of a noun lemma. Since reviewers raised concerns about the validity of the class-based measure used in the confirmatory analysis (PND class), this exploratory analysis serves the purpose of testing the robustness of

Table 7
Results of Bayesian analysis for pooled data from all the languages. The estimated mean is reported as well as the lower and upper limits of the 95% credible interval and the probability *P* of the effect being < 0 (for negative coefficients) or > 0 (for positive coefficients), given the data. Boldface indicate strong evidence (Bayesian P ≥ 0.95).

Comparison	Mean	Lower	Upper	*P*
Intercept	2.16	1.77	2.56	1
PND class	**1.87**	**0.64**	**3.06**	**1**
Surface-form frequency	**0.49**	**0.29**	**0.69**	**1**
Finnish vs Estonian	−0.34	−0.81	0.11	.93
Polish vs Finnish & Estonian	**−0.32**	**−0.61**	**−0.04**	**.99**
Age	**0.67**	**0.43**	**0.92**	**1**
PND class:Surface-form frequency	**−0.63**	**−1.37**	**0.12**	**.95**
PND class:Finnish vs Estonian	**−1.41**	**−2.47**	**−0.36**	**1**
PND class:Polish vs Finnish & Estonian	0.02	−1.05	1.06	.52
Surface-form frequency:Finnish vs Estonian	**−0.19**	**−0.41**	**0.03**	**.96**
Surface-form frequency:Polish vs Finnish & Estonian	**0.14**	**−0.02**	**0.29**	**.96**
Age:Finnish vs Estonian	−0.22	−0.56	0.1	.90
Age:Polish vs Finnish & Estonian	−0.02	−0.17	0.13	.63
PND class:Age	−0.09	−0.32	0.16	.77
Surface-form frequency:Age	−0.01	−0.16	0.14	.56
PND class:Surface-form frequency:Finnish vs Estonian	0.27	−0.41	0.96	.79
PND class:Surface-form frequency:Polish vs Finnish & Estonian	−0.32	−0.95	0.32	.84

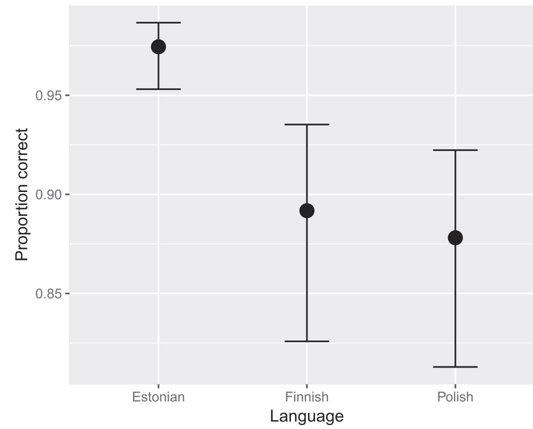


Fig. 9. Proportion correct overall in all three languages.

The PND form measure is described in Appendix C.1. In brief, the measure defines neighbourhood on the basis of the phonemic *transformations required to derive a particular target form* from the base form (nominative). The measure is therefore sensitive to any transformations in the stem or the affix on the individual-form level. The particular PND form value of each form represents not just the number of its neighbours but also the ratio of the summed similarities of neighbours with

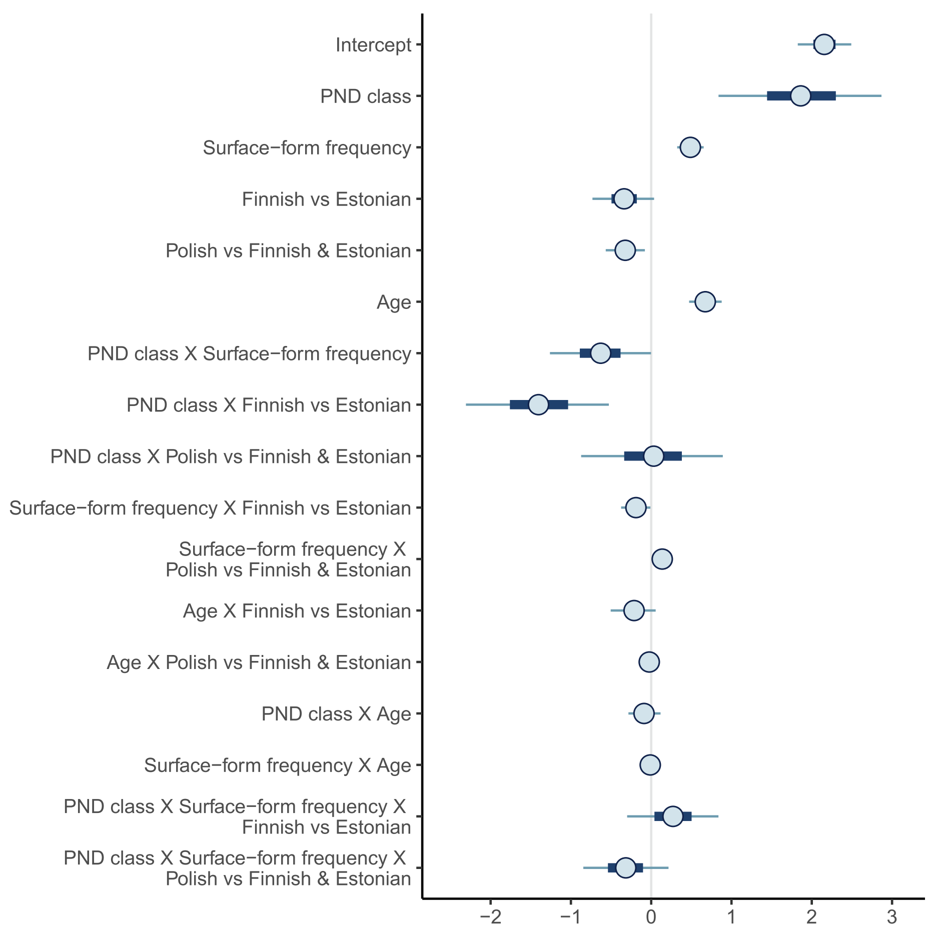


Fig. 8. Bayesian mean estimates and 95% credible intervals for all predictors in the pooled model.

findings across different operationalisations of PND, as well as differentiating further between the theories: Recall from the Introduction that an effect of PND form is predicted by connectionist and exemplar accounts, but is difficult to explain under rule/class-based accounts.

the target form normalised by the summed similarities of items not considered neighbours. Hence, the final value is representative of the *strength* of the neighbourhood for successful analogy as well as the level of *interference* from similar non-neighbours or “enemies”. The analysis

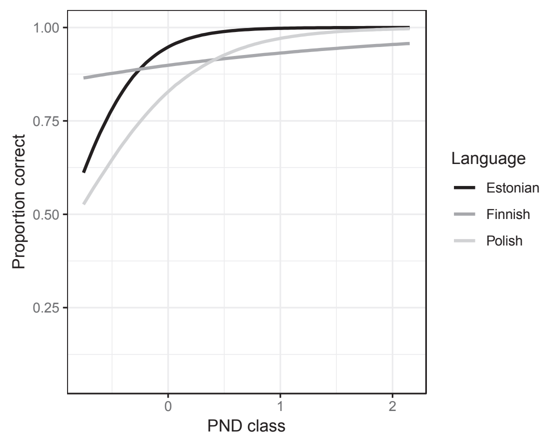


Fig. 10. Pooled Bayesian analysis: effect of log PND class in the three languages.

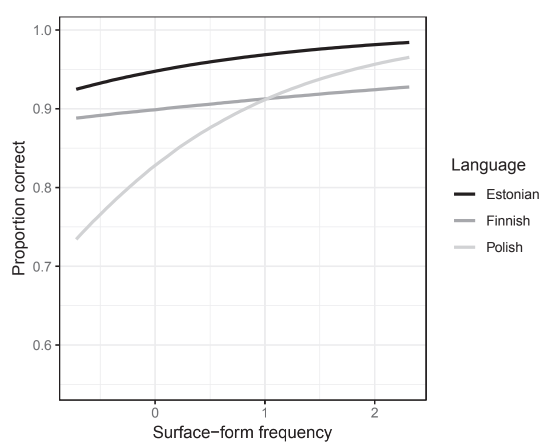


Fig. 11. Pooled Bayesian analysis: effect of log surface-form frequency in the three languages.

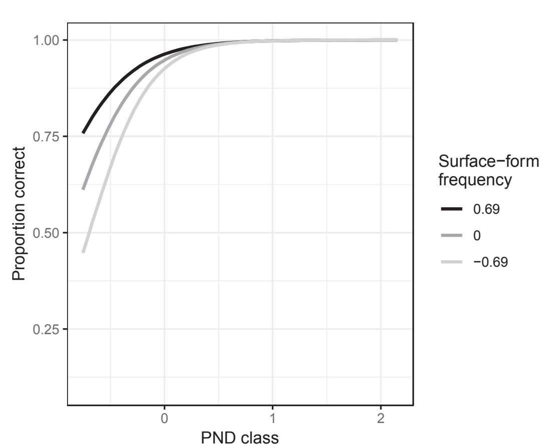


Fig. 12. Bayesian pooled analysis: effect of log PND class as conditioned on different values of surface-form frequency (mean ± 1SD).

was performed using Bayesian models only. Detailed results can be found in Appendix C.2. Here, we present a summary in Table 9, where '*' indicates results differing from the confirmatory results listed in Table 8.

As Table 9 shows, all findings reported in the confirmatory analyses using the PND class measure were supported with the new PND form measure. In addition, effects of surface-form frequency, PND and their interaction were found with stronger evidence when using PND form as opposed to PND class. Under the assumption that the finer-grained PND measure is more sensitive to actual effects of phonological analogy, these results provide support for pervasive effects of PND and surface-form frequency in all three languages. They also provide further support for constructivist and exemplar over rule-based accounts, since the latter predict only class-based effects of PND. Finally, the PND form analysis found strong evidence for an interaction between PND and age in Polish ($P = 0.95$), and also some evidence for it in Finnish ($P = 0.91$). The interaction suggests that the effect of PND becomes smaller with increasing age, possibly reflecting the development of more abstract representations.

Descriptive error analysis

In this section we examine the errors made by children in our experiment. As we found effects of both surface-form frequency and PND (both PND class and PND form) in the analyses reported above, we are interested in assessing whether the types of errors they made also reflect these factors.

Children may use more frequent forms of a noun in place of less frequent forms. This type of error shows an effect of surface-form frequency resulting in errors in inflection. Children are also known to make errors of **overgeneralization**. For the present analysis, we defined overgeneralization as the use of an inflection that is of the correct target case (e.g., genitive) but is not the correct inflection for that target noun (e.g., a genitive marker from a different declension class). For example, *telefon*, 'telephone' takes the genitive form *telefonu*, but children sometimes overgeneralize the (more-frequent) genitive marker $-a$ yielding **telefona* (perhaps by analogy with forms such as *smok* → *smoka*, 'dragon'). Since Polish has different declensions for all cases, these errors are possible, in principle, for all target nouns. For Finnish, different declension forms are available only for partitive and illative targets, and for Estonian only partitive and genitive targets. Errors like these, across declension classes, may reflect the influence of PND, whereby children overgeneralize an ending from a larger neighbourhood, whether defined across the whole paradigm (PND class), or with respect to the target form (PND form). Some overgeneralization errors may also reflect the use of a more frequent target form, reflecting the influence of surface-form frequency. This section gives a descriptive overview of errors made by the children in each of the languages in our study.

Polish

In Polish, 19% of all scorable responses contained errors (total: 570 errors). Table 10 shows the most common types of errors made by the Polish participants. Most often, Polish children produced nominative case instead of the target case. In some classes, where the nominative and accusative forms are syncretic, this means using either the nominative or accusative form of the noun. This can be explained as a surface-form-frequency error: in Polish child-directed speech corpora, nominative and accusative forms of each noun have the highest surface-form frequency. Several error types that occurred in Polish can also be explained by overgeneralisation, such as using the genitive $+a$ affix instead of genitive $+u$, or dative $+owi$ affix instead of dative $+u$ (in both instances, children were choosing an affix that occurs in more classes than the target affix). The use of the locative affix $+e$ instead of locative $+i$ in can be explained either as a surface-form-frequency error or by overgeneralisation. The other most common errors cannot be explained in this way. Most commonly, children shortened the stem, produced a less frequently occurring form of the noun or added an affix from a less frequent noun class (sometimes resulting in unattested forms of the noun).

Finnish and Estonian

In Finnish, 18.7% of scorable responses involved errors (total 530 errors), while in Estonian, 8.5% of all scorable responses constituted

Table 8
Summary of observed effects in confirmatory (by language) and pooled analyses with respect to each hypothesis. Checkmark indicates effect supported in both analyses (significant in frequentist analysis *and* strong evidence in the Bayesian analysis, i.e., $P \geq 0.95$); parentheses indicate effect found in only one of the two analyses.

	Age	Surface-form frequency	PND class	Surface-form frequency*PND class	Surface-form frequency*Age	PND class*Age
Polish	✓	✓	(✓)	(✓)	✗	✗
Finnish	✓	✗	(✓)	✗	✗	✗
Estonian	✓	✓	✓	✗	✗	✗
Pooled	✓	✓	✓	✓	✗	✗

Table 9
Summary of observed effects in exploratory Bayesian analysis with PND form with respect to each hypothesis. Differences with the confirmatory analysis are indicated by '*'. Parentheses indicate where an effect was found with *weak* evidence only, i.e., $0.9 \leq P \leq 0.95$.

	Age	Surface-form frequency	PND form	Surface-form frequency*PND form	Surface-form frequency*Age	PND form*Age
Polish	✓	✓	✓*	✓*	✗	✓*
Finnish	✓	(✓)*	✓*	✗	✗	(✓)*
Estonian	✓	✓	✓	✗	✗	✗

errors (total 245 errors). For coding the errors in Finnish and Estonian, a distinction between 'stem errors', 'case affix errors' and 'class affix errors' was made in order to group types of erroneous responses. *Stem errors* occurred only in the stem of a noun form. *Case affix errors* involve affixes that are never used to mark the target case, with any noun in the language. Thus, both *+A* and *+(t)tA* are considered correct case affixes for partitive in Finnish, regardless of the noun class, but any other case affix used with a partitive target was coded as a case affix error. However, *class affix errors* involve erroneous affix selection for the noun declension class in question, i.e. the use of the *+A* affix rather than the *+(t)tA* affix in partitive case (e.g., production of *hame+a* instead of target *hame+tta* for 'skirt' in partitive case in Finnish). Recall that these class affix errors are only applicable in partitive and illative case in Finnish, and partitive and genitive case in Estonian, in which variant endings occur, depending on declension class.

Tables 11a and b display the most frequent types of errors produced by the Finnish and Estonian participants. The most frequent by far, in both languages, was stem errors (66% of all errors in Finnish and 76% in Estonian), followed by errors involving an incorrect case affix (29% of errors in Finnish; 17% in Estonian), or an erroneous class affix (12% of all errors in Finnish and 15% in Estonian). Note that some erroneous responses include both stem and affix errors; hence, percentages add up to more than 100%.

In both languages, many frequently-occurring errors can be attributed to overgeneralization. The single most frequent error type in both languages was the omission of stem changes, meaning that children attached the case affix to the nominative stem form, which is both the most frequent and the simplest stem form. Some errors in both languages can be explained by children using a more frequently occurring form of the noun other than the nominative – these were case affix errors, in which a more frequently occurring case affix was used. However, several other errors cannot be explained by either overgeneralization or form-frequency effects. Children sometimes used the inflectional paradigm of a smaller noun class than that of the target noun, or a case form lower in frequency than the target form. A possibility here is that some errors may be driven by language-specific factors, as explored below.

It should be mentioned that the unscorable responses included some response types that might be considered "escape strategies". Unscorable responses included a sizable number of pre- and postpositional phrases (12.2% in Polish, 11.7% in Estonian and 2.9% in Finnish; see Table 3) and substitutions of partitive for genitive targets (3.1%, only in Estonian). All of these are grammatical in the language and so could not be considered errors, but neither could we score them as correct for the targeted case; hence, we omitted them from the analysis. It is worth noting, however, that 68% of the partitive for genitive responses in Estonian were in the three classes with stem-changing consonant gradation (class VI, VII and VIII), and 46% were for nouns in class VII, in which nominative and partitive are syncretic forms. Hence, the selection of a partitive form despite training with genitive responses may reflect an avoidance of stem-changing forms, as well as use of forms with higher frequency.

Summary

The analysis of children's errors demonstrated that both surface-form-frequency errors and overgeneralization errors occur in the production of case marking in all three languages. However, in all languages, there were also errors that could not be accounted for by either of these factors.

Table 10
The most frequently occurring errors in Polish, along with target case, number of occurrences of each error (N) and proportion of that type of error out of all errors (%). In the 'Example' column, the word in parentheses represents the target form. The final column ('Error classification') shows which errors can be attributed to the use of a more frequent form of a noun ('frequency') or to the use of overgeneralisation. Only errors in which $N \geq 10$ are shown.

Error	Target case	*N*	%	Example	Error classification
Nominative case	all cases	101	17.7	*grzebień (grzebieni+owi)*	frequency
+a instead of Ø	Accusative	91	16.0	*traktor+a (traktor)*	
+a instead of *+u*	Genitive	63	11.1	*telefon+a (telefon+u)*	overgeneralisation
+u instead of *+e*	Locative	42	7.4	*aparat+u (aparaci+e)*	
+u instead of *+owi*	Dative	28	4.9	*pociąg+u (pociąg+owi)*	
+owi instead of *+u*	Dative	27	4.7	*koł+owi (koł+u)*	overgeneralisation
shortened stem	all cases	13	2.3	*wielob+owi (wieloryb+owi)*	
+e instead of *+i*	Locative	12	2.1	*koszul+e (koszul+i)*	frequency/overgeneralisation
+a instead of *+owi*	Dative	12	2.1	*aparat+a (aparat+owi)*	
+u instead of *+a*	Genitive	11	1.9	*traktor+u (traktor+a)*	
Total		387	70.7		

Table 11
The most frequently occurring errors in Finnish (a) and Estonian (b), with number of occurrences (N), proportion of each type of error out of all errors (%) and examples. In the example column, the target form is given in parentheses (along with the nominative form, in the case of stem errors). The final column ('Error classification') shows which errors can be attributed to the use of a more frequent form of a noun ('frequency') or overgeneralisation. Only errors in which $N \geq 10$ are included in the tables. As some erroneous responses include both stem and affix errors, the sum of types of error found in Finnish and Estonian amounts to more than 100%.

Error	Type	*N*	%	Example	Error classification
(a) Finnish					
No stem change	stem	160	30.2	*kuusi + n* (*kuuse + n*, nominative: *kuusi*)	overgeneralisation
Class XI: stem -*s* > -*kse*	stem	64	12.1	*rengakse + n* (*renkaa + n*, nominative: *rengas*)	
Shortened stem	stem	59	11.1	*kauli + ssa* (*kaulime + ssa*, nominative: *kaulin*)	
Lengthened stem	stem	46	8.7	*kaarime + sta* (*kaare + sta*, nominative: *kaari*)	
+ *A* in partitive	class	34	5.7	*kone + a* (*kone + tta*, partitive)	overgeneralisation
Nominative case	case	29	5.5	*puhelin* (*puhelime + en*, illative)	frequency
Illative case	case	25	4.7	*maalaukse + en* (*maalaukse + lle*, allative)	
Allative case	case	24	4.5	*käpy + lle* (*käpy + yn*, illative)	frequency
Genitive case	case	24	4.5	*hai + n* (*hai + hin*, illative)	frequency
Elative case	case	21	4.0	*uuni + sta* (*uuni + in*, illative)	
+ *Vn* in Illative	class	18	3.0	*aarre + en* (*aartee + seen*, illative)	overgeneralisation
Partitive case	case	13	2.5	*kone + tta* (*konee + sta*, elative)	frequency
Total		517	94.0		
(b) Estonian					
No stem change	stem	120	49.0	*päike + ga* (*päikese + ga*, nominative: *päike*)	overgeneralisation
+ *V* instead of gradation	stem	39	15.9	*sulge + le* (*sule + le*, nominative: *sulg*)	overgeneralisation
Partitive + *t* instead of Ø	class	21	8.6	*kruvi + t* (*kruvi +* Ø, partitive)	overgeneralisation
Nominative case	case	11	4.5	*masin* (*masina + le*, allative)	frequency
lengthened stem	stem	11	4.5	*patareisi + le* (*patarei + le*, nominative: *patarei*)	
Allative case	case	10	4.1	*madratsi + le* (*madratsi + st*, elative)	frequency
Total		212	86.5		

Exploratory analyses of language-specific features affecting performance

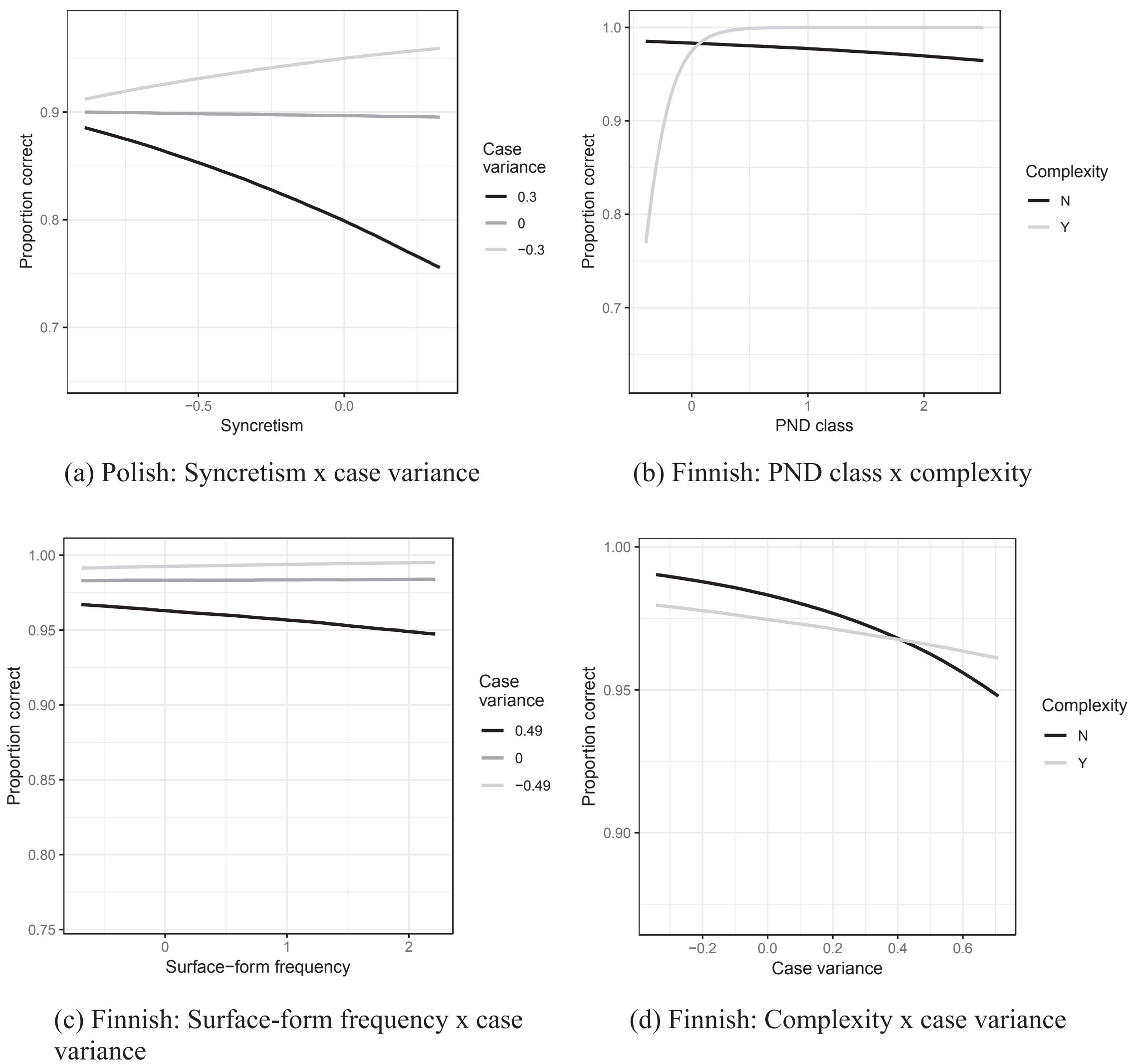
The analyses reported above – both the confirmatory analyses set out in our preregistration and the non-preregistered all-languages and PND-form analyses – were designed to mediate between different theoretical approaches to morphological acquisition. The final set of analyses reported below have a very different goal; simply to explore additional language-specific phenomena that may occur within the domain of morphological acquisition, and that – if confirmed in subsequent preregistered studies – would need to be explained by future theoretical accounts. A secondary goal is to investigate whether the findings reported above hold after controlling for these language-specific factors. It is important to stress that these analyses were designed after seeing the data and are therefore subject to "researcher degrees of freedom" with regard to decisions such as how to code the new, exploratory predictors.

These analyses explore the possible effects of three linguistic predictors on the accuracy of responses:

- **Case variance** (all languages, continuous predictor): The number of different affixes marking the case of the target form (e.g., gentive); i.e. the number of different morphological markers that a participant must "choose" between on a given trial. For example, if the target on a particular trial is the Polish genitive form *mroku,* 'darkness', the value for case variance is 4, since, in addition to -*u*, Polish marks the genitive with -*a*, -*y* and -*i*. We would anticipate a negative effect of case variance (i.e., the greater the number of affixes to "choose" between on a given trial, the greater the likelihood of error).
- **Affix syncretism** (Polish only, Finnish and Estonian do not exhibit affix syncretism, continuous predictor): The number of cells in the paradigm filled by the affix that is the target for that particular trial. For example, if the target on a particular trial is the Polish genitive form *mroku,* 'darkness', the value for affix syncretism is 5, since, in addition to masculine genitive, -*u* marks neuter dative, neuter locative, masculine locative and masculine vocative. The expected direction of this effect is not entirely clear. On the one hand, a high degree of affix syncretism could be expected to cause a degree of confusion. On the other hand, if a particular affix fills a large number of cells, participants may sometimes produce the correct target form fortuitously, even if they mistake the case and gender required by the context.
- **Stem-change** (Finnish and Estonian only, stem change is not a criterion used to define Polish classes and few of our experimental stimuli exhibited stem change, binary predictor): Whether, for a given trial, the stem of the target form is different (1) or the same (0) as the stem of the normative (citation) form of that noun. For example, if the target on a particular trial is the Estonian genitive form *hobuse*, 'horse', the value for stem-change would be 1, since the nominative form is *hobune* (i.e. the stem changes from *n* to *s* and -*e* is added). However, if the target on a particular trial is the Estonian partitive form *kuud*, 'moon', the value for stem-change would be 0, since the nominative form is *kuu* (i.e. the stem does not change before -*d* is added).

These factors were chosen as potentially relevant on the basis of our experience with the languages, informal inspection of corpus data, and previous studies (e.g. Argus, 2009b; Laalo, 2009; Laaha & Dressler, 2012; Kjærbæk et al., 2014; Slobin, 1985). Furthermore, the three languages in our study differ in relevant ways across these three factors. We did not intend this set of factors to be exhaustive; neither did we select them on the basis of a particular theoretical approach, but rather as potentially cognitively relevant and typologically variable predictors.

Bayesian models were used exclusively for these exploratory analyses due to multiple convergence issues with the equivalent frequentist models. We began with the same binomial models as for the main analysis, and added the predictors set out above for each language, with all continuous predictors scaled and centred. We included all two-way interactions in the model, but no three-way or higher interactions, as previous analyses showed a lack of power. Maximal models were run for each language, with random intercepts for participant and noun and by-participant random slopes for all fixed effects.



(a) Polish: Syncretism x case variance

(b) Finnish: PND class x complexity

(c) Finnish: Surface-form frequency x case variance

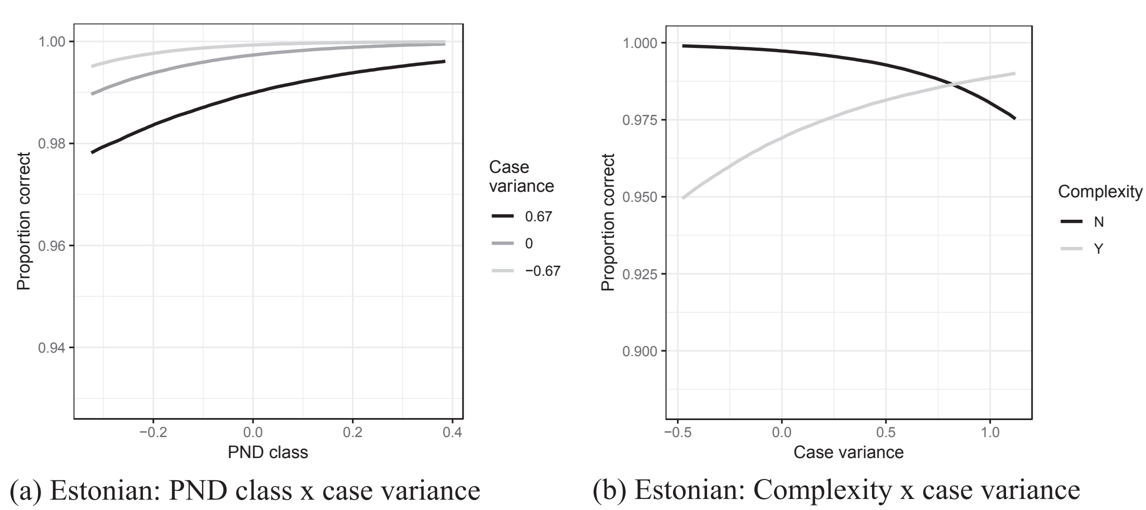
(d) Finnish: Complexity x case variance

Fig. 13. Interaction effects with strong evidence in exploratory analysis for Polish and Finnish.

Results of exploratory analyses

The detailed results of the exploratory analyses for each language can be found in Table E.1 in Appendix E. Figs. 13 and 14 illustrate the interactions that were supported by strong evidence.

For Polish, the only main effect with strong evidence was case variance, indicating that children's accuracy was lower for cases with highly variable suffixes. There was strong evidence only for an interaction between syncretism and case variance: the effect of syncretism



(a) Estonian: PND class x case variance

(b) Estonian: Complexity x case variance

Fig. 14. Interaction effects with strong evidence in exploratory analysis for Estonian.

became increasingly negative for higher case variance (i.e., highly syncretic affixes were more detrimental to accuracy in cases with variable affixes — see Fig. 13a). Perhaps unsurprisingly, the addition of the new predictors decreased the strength of the evidence for the previously-observed effects of PND class ($P = 0.91$), surface form frequency ($P = 0.89$) and their interaction ($P = 0.86$). However, although these effects no longer meet our criteria for “strong evidence”, these P_{MCMC} values still constitute greater evidence for these effects than for their absence.

For Finnish, the model revealed a strong, positive main effect of PND class, but not surface frequency (as in the main analysis). It also revealed strong evidence for a negative effect of case variance, with children showing less accuracy on cases with more variant forms. Three interactions were observed with strong evidence: PND class interacted with stem-change, such that the facilitating effect of phonological neighbourhood density on children’s accuracy was greater for forms with a stem change (see Fig. 13b). Case variance interacted with surface-form frequency and stem-change: The negative effect of high case variance on accuracy was more pronounced for highly frequent forms, contrary to our expectations, and also more pronounced for cases with variable affixes (see Fig. 13c and d).

For Estonian, the model revealed strong positive effects of PND class and surface-form frequency (again, as in the main analysis). It also revealed strong evidence for a negative effect of stem-change as well as some evidence ($P = 0.91$) for a negative effect of case variance, suggesting that participants’ performance was worse for targets with a stem change or in cases with high inflectional variation (partitive and genitive). The effect of case variance interacted with PND class and with stem change: The positive effect of PND class was smaller for cases with variable inflections, while the negative effect of stem change was larger for cases with variable inflections (see Fig. 14a and b).

Summary

Overall, the exploratory analyses suggest that, in addition to the language-general factors investigated in the main analysis, children’s accuracy in case marking was influenced by language-specific factors. In particular, in all three languages, variance in the affixes available to express a particular case (high case variance) negatively impacted on children’s acquisition. Stem-changes negatively impacted acquisition for Estonian and, for Finnish, seemed to cause a greater reliance on phonological analogy (PND class). That said, given that these analyses were merely exploratory, these apparent effects must await confirmation from preregistered studies. Importantly, all effects of language-general factors observed in the main confirmatory analyses held after controlling for these exploratory language-specific factors, though they were somewhat weakened for Polish; for Finnish and Estonian, they were essentially unchanged.

Discussion

The present investigation constituted a large-scale, preregistered, elicited-production study of noun case marking, with the same design (as far as possible) across three languages. The aim was to identify which main effects and interactions previously posited in the literature (but only sometimes observed) would hold across languages, and which represent genuine cross-linguistic differences, as well as to investigate – in a purely exploratory fashion – further language-specific factors that have received little experimental, cross-linguistic attention.

Our confirmatory, pre-registered analyses found effects of surface-form frequency for Polish and Estonian, and a class-based measure of phonological neighbourhood density (PND class) for all three languages. Individually, only Polish showed evidence of a predicted interaction, such that the effect of PND class was greater for forms with lower surface-form frequency, which are less available for direct retrieval from memory. An exploratory all-languages analysis yielded evidence for both main effects and the two-way interaction, but no three-way interaction with language (suggesting that the failure to observe an interaction of PND class by surface-form frequency for Finnish and Estonian may be due to low power). Exploratory language-by-language analyses using a form-based measure of PND (PND form) not only replicated the pattern set out above, but yielded stronger evidence for each effect. This analysis also yielded strong evidence for an interaction between age and PND for Polish ($P = 0.95$) and somewhat weaker evidence for the same interaction for Finnish ($P = 0.91$). In both cases, the interaction was negative, suggesting that, counter to our predictions, neighbourhood density becomes less important with age. Interestingly, this negative interaction has two possible explanations that are almost polar opposites: Phonological analogy may become less important with age either because development involves (a) building more abstract knowledge of the system or (b) increasing storage of ready-inflected surface forms, both of which would render phonological analogy superfluous. However, this finding should be treated as tentative, given that it was not predicted, and holds only using the exploratory, form-based measure of phonological neighbourhood density (PND form, but not PND class). Indeed, although all languages showed a main effect of age (i.e., older children unsurprisingly made fewer errors), developmental changes in the importance of surface-form-frequency and PND were, in general, not observed (though, again, we cannot rule out low power as an explanation).

Returning to the findings for which strong evidence was observed, we now turn to the question of the implication of these findings for the theoretical models outlined in the Introduction. (We do not consider here the exploratory, language-specific analyses, as they were not designed to mediate between different theoretical accounts; other than to emphasize that the theoretically-relevant effects observed were generally robust to the inclusion in the statistical models of these language-specific factors).

First, the finding that an effect of surface-form frequency was observed for Polish, Estonian and all languages together constitutes support for proposals towards the similarity-based end of the continuum. As we noted in the Introduction, while advocates of rule-based theories do not explicitly predict the *absence* of such effects, neither do they – unlike similarity-based models – explicitly build them into the acquisition mechanism; at best, explaining them as a processing-based add-on. Second, the finding of – for every language – an effect of class-based phonological neighbourhod density (PND class) is compatible, in principle, with all approaches. That is, the finding is compatible with both rule-based accounts that posit explicitly-represented declension classes, and with similarity-based accounts which represent such classes implicitly in the form of patterns of hidden-unit activation or analogical similarity. However, the finding of an effect of form-based phonological neighbourhood density (PND form) is compatible only with connectionist and exemplar approaches since, by definition, this measure not only cuts across discrete declension classes, but does so in a way that differs depending on the target form.

A more equivocal finding was the interaction of surface-form frequency and phonological neighbourhood density, which (whether in its class- or form-based variant) was observed only for Polish individually (though also, in an exploratory analysis, when collapsing across all languages). Again, this finding – if subsequently confirmed – is difficult to explain under rule-based accounts, which posit no mechanism that would yield such an effect, but it falls naturally out of connectionist and exemplar accounts: Recall from the Introduction that connectionist models predict this interaction (and observe it; e.g., Engelmann et al., 2019), because high-frequency pairs (e.g., stem-output) develop their own dedicated input–output mappings, while low-frequency pairs rely more on the abstract representations stored in the hidden units. Exemplar models predict this interaction because phonological analogy is necessary only for lower-frequency (or novel) exemplars that are less likely to be retrieved directly. Note that, just as for the main effect of PND form, the multiple-rules, pre-/proto-morphological and words and rules approaches struggle to explain the interaction of surface-form frequency by PND form, since they incorporate only a class-based notion of phonological neighbourhood density.

Indeed, although exploratory, we would suggest that our new connectionist-inspired measure of PND form represents one of the most important novel contributions of the present work, as it addresses a long-standing problem for class-based measures: the fact that the declension (or conjugation) class of a noun (or verb) effectively changes, depending on the particular target that the speaker is intending to produce. In general, class- and rule-based accounts are well suited to capturing highly regular systems (e.g., *[VERB]-ed*), but not the many cross-cutting subregularities found in complex systems of noun case marking. For example, recall that reference grammars for Finnish and Estonian posit in the region of 50 separate noun-declension classes defined by phenomena such as inflectional stem-final vowel variation and syllable gradation. For such systems, an account based on classes, rules or schemas faces an intractable lumping or splitting dilemma. At one extreme (lumping), it can posit a single class/rule/schema, which does not account for any of the fine-grained phonological differences between them. At the other extreme (splitting), it can posit around 50 separate classes/rules/schemas, which does not account for the many overarching similarities between them. Any intermediate position is an unhappy compromise that fails to capture *both* differences and similarities. Form-based, connectionist and exemplar approaches sidestep the problem by dispensing with classes, rules and schemas altogether. Indeed, if the utility of a form-based measure of phonological neighbourhood density is confirmed by subsequent preregistered studies, domains beyond acquisition – for example language evolution (e.g., Dale & Lupyan, 2012; Lupyan & Dale, 2010) – could potentially benefit from its adoption (we thank an anonymous reviewer for raising this point).

Can we choose between a connectionist and an exemplar approach? A complication here is that, within each framework, a wide variety of different implementations are possible. For example, a smaller number of hidden units – or a stronger analogical mechanism – will increase the model's reliance on support from phonological friends or neighbours (i.e., the effect of PND), at the expense of stored tokens (i.e., the effect of surface-form frequency). Furthermore, the impact of such a change will depend on the system being learned: It would be catastrophic for a system with a large proportion of unpredictable irregulars, but not for a system with a small number of phonologically-consistent families. Thus, the data from the present study cannot be used to differentiate between exemplar and connectionist approaches in general — only between particular computational instantiations of these accounts. This is a project that we are currently undertaking (e.g., Engelmann et al., 2018), but is clearly beyond the scope of the present paper.

In the meantime, one theoretical advantage of connectionist approaches is that, while exemplar approaches tend to be static, operating on the basis of an adult-like system, connectionist approaches are typically more dynamic, and are hence able to explain developmental changes (e.g., Elman, Bates, & Johnson, 1998; Westermann & Ruh, 2012). Hence, they are potentially better placed to explain the present finding that, at least for Polish, the effect of phonological neighbourhood density appears to decrease with age.

In conclusion, while it remains for future computational-modeling work to mediate between different similarity-based approaches, the findings of the present study make an important theoretical contribution in pointing away from rule-based theoretical accounts of morphological acquisition, and towards the similarity-based end of the continuum. More rule-based accounts struggle to explain the observed effects of surface-form frequency and form-based phonological neighbourhood density; findings that any successful account of the acquisition of inflectional morphology will need to explain.

Conflict of interest

The authors declared that there is no conflict of interest.

Acknowledgements

We are grateful to two anonymous reviewers and Vera Kempe, whose comments helped greatly improve this manuscript. We thank all the nurseries and children for participating in the study. Many thanks to Johanna Kahur and Pille Markov, who helped collect data in Estonia, and to Minna Kirjavainen, who collected and curated the Kirjavainen-Max Planck Finnish corpus. Thank you also to Dr Marta Szreder for her help with the classification of nouns in Polish. This work was funded by the UK Economic and Social Research Council (ESRC, grant number: ES/L008955/1). The Estonian component was funded by a grant from the European Union's Seventh Framework Programme (Marie Curie IEF grant no 623742 to V.-A. Vihman) and support from the Centre of Excellence in Estonian Studies (European Union, European Regional Development Fund).

Appendix A. Description of noun classes in Polish, Finnish and Estonian

According to the most widely accepted noun classification by Tokarski (1973), Polish has 17 declension classes (5 masculine, 6 feminine and 6 neuter), based solely on gender and phonological features. However, this classification has often been modified (cf. Gruszczyński & Bralczyk, 2002). For the purpose of this study, we classified nouns according to phonological and morphosemantic criteria (gender, animacy and countability), which resulted in 12 declension classes (for detailed information, see Table A.2).

Polish distinguishes three grammatical genders, and the gender of most nouns can be predicted from the phonological form of the nominative: nouns ending in +*a* are usually feminine, nouns ending in +*o*, +*e*, or +*ę* are neuter, nouns which end in a 'hard' (unpalatalised) consonant are masculine. 'Soft' (palatalised) consonants can be either masculine or feminine. Within the system, the same case may be signalled by different affixes

Table A.1
Summary of the 7 Polish noun cases, including singular case ending, function and type frequency (%, for singular nouns in the CDS corpora, interpolated for syncretic forms). The final column indicates the cases tested in the current study. The ['] symbol before an affix indicates palatalisation of the preceding consonant(s).

Case	Affix (gender)			Main usage	%	
	Feminine	Masculine	Neuter			
Nominative	+*a*/Ø	Ø	+*o*/+*e*	Citation form, subject	26.7	✓
Genitive	+*y*/+*i*	+*u*/+*a*	+*a*	Direct object of certain verbs, negation	19.9	✓
Dative	+*y*/+*i*/+'*e*	+*owi*	+*u*	Indirect object, experiencer	6.6	✓
Accusative	+*ę*/+Ø	+*a*/Ø	+*o*/+*e*	Direct object, direct object of certain prepositional phrases	22.6	✓
Instrumental	+*ą*	+*em*/+'*em*	+*em*/+'*em*	Subject predicative, direct object of certain verbs	8.8	✓
Locative	+'*e*/+*i*/+*y*	+*u*/+'*e*	+*u*/+'*e*	Direct object of certain prepositional phrases	11.8	✓
Vocative	+*o*/+*i*	+*u*/+'*e*	+*o*/+*e*	Form of address	3.7	

Table A.2
Summary of the 12 Polish declension classes for singular nouns, along with stem coda (soft/hard consonant or *k*-), animacy ('Anim.') and countability ('Count.'), applies only to the masculine classes), case formation, and CDS corpus type frequency (in %). The final column indicates which classes were included in the current study. The [‘] symbol before an affix indicates palatalisation of the preceding consonant(s).

Class	Stem coda	Animate	Count.	Nominative	Genitive	Dative	Accusative	Instrumental	Locative	Type %	
Feminine1	*soft consonant-*	NA	NA	*+a*	*+y/+i*	*+y/+i*	*+ę*	*+ą*	*+y/+i*	7.7	✓
Feminine2	*hard consonant-/k-*	NA	NA	*+a*	*+y/+i*	*+‘e*	*+ę*	*+ą*	*+‘e*	30.4	✓
Feminine3	*soft consonant-*	NA	NA	Ø	*+y/+i*	*+y/+i*	Ø	*+ą*	*+y/+i*	4.5	
Masculine1	*soft consonant-/k-*	no	no	Ø	*+u*	*+owi*	Ø	*+‘em*	*+u*	7.1	✓
Masculine2	*hard consonant-*	no	no	Ø	*+u*	*+owi*	Ø	*+em*	*+‘e*	12.8	✓
Masculine3	*soft consonant-/k-*	no	yes	Ø	*+a*	*+owi*	Ø	*+‘em*	*+u*	9.5	✓
Masculine4	*hard consonant-/k-*	no	yes	Ø	*+a*	*+owi*	Ø	*+em*	*+’e*	2.1	✓
Masculine5	*soft consonant-/k-*	yes	NA	Ø	*+a*	*+owi*	*+a*	*+‘em*	*+u*	9.9	✓
Masculine6	*hard consonant-*	yes	NA	Ø	*+a*	*+owi*	*+a*	*+em*	*+‘e*	4.5	✓
Neuter1	*soft consonant-/k-*	NA	NA	*+o*	*+a*	*+u*	*+o*	*+’em*	*+u*	3.6	✓
Neuter2	*hard consonant-*	NA	NA	*+o*	*+a*	*+u*	*+o*	*+em*	*+‘e*	3.0	✓
Neuter3	*hard consonant-*	NA	NA	*+e*	*+a*	*+u*	*+e*	*+’em*	*+u*	4.9	

Table A.3
Summary of the 14 Finnish and Estonian cases, including singular case ending, function ('Main usage') and type frequency (%, calculated over singular nouns and adjectives in the CDS corpora, and interpolated for syncretic forms in Estonian). The final column indicates cases tested in the current study; check marks in parentheses show cases which were tested in one of the two languages. Capital letters in the affix columns denote variable realisation depending on the context. 'V' represents a variable vowel, determined differently in the two languages (in Finnish, a vowel identical to the preceding vowel; in Estonian, a lexically determined vowel (*a, e, i* or *u*), usually unpredictable from phonological context). 'A' signifies variation according to vowel harmony (*a/ä*).

Case	Singular affix Finnish	Singular affix Estonian	% Finnish	% Estonian	Main usage	
Nominative	Ø	Ø	33.5	26.2	Subject, citation form	
Genitive	*+n*	*Ø/+V*	16.3	20	Total object, possession, object of postposition	✓
Partitive	*+(t)tA/+A*	*Ø/+V/+t/+d*	16.9	18	Partial object, negation	✓
Inessive	*+ssA*	*+s*	6.7	4.6	'Inside'	✓
Illative	*+(h)Vn/+seen*	*+sse*/short form	7.2	10.7	'Into'	(✓)
Elative	*+stA*	*+st*	5.8	4.2	'Out of'	✓
Adessive	*+llA*	*+l*	5.4	3.5	'On', 'at', experiencer	
Allative	*+lle*	*+le*	4.1	3.8	'To', recipient	✓
Ablative	*+l*	*+lt*	1.9	0.3	'(Away) from'	
Comitative	[plural only]	*+ga*	–	7.2	'With'	(✓)
Abessive	*+ttA*	*+ta*	0	0.3	'Without'	
Essive	*+nA*	*+na*	0.6	0	Temporary state	
Translative	*+ksi*	*+ks*	11.5	1.1	Change of state	
Instructive (Finnish)	[plural only]	–	–	–	Means used	
Terminative (Estonian)	–	*+ni*	–	0	'Until', 'up to'	

across classes, and the same affix may mark different cases, even within the same class. Some of these affix choices are systematic – for example, in the locative case, the choice of affix depends on the phonological properties of the final consonant of the stem; in the accusative, on animacy; and in the genitive, on a combination of semantic, morphological and lexical factors. However, semantic cues are not entirely reliable and thus any

Table A.4
Summary of the 17 Finnish declension classes for singular nominals, along with the number of stems, partitive and illative case formation, final phoneme(s), any stem changes, and CDS corpus type frequency (%, including nouns and adjectives). 'V' denotes a vowel which is identical to the preceding vowel, and other capital letters signify variation according to vowel harmony. The final column indicates the classes included in the current study. Class III contains mostly monosyllables. Classes I and IX are considered to be the most productive classes. *Class IX nouns often end in *−Us*, but also *−Vs*, while class XI nouns mostly end in *−As*, but also *−es* and *−is*.

Class	Stems	Partitive	Illative	Final	Stem changes	Type %	
I	1	*+A*	*+Vn*	short *V*	none	75.2	✓
II	1	*+tA*	*+Vn*	V_1V_2	none	1.0	
III	1	*+tA*	*+hVn*	*VV*(monosyllable)	none	1.7	✓
IV	1	*+tA*	*+seen*	V_1V_1	none	0.2	
V	2	*+A*	*+Vn*	short *i*	*−i > −e*	2.7	✓
VI	2	*+tA*	*+Vn*	*r/l/n*	*+e*	0.4	
VII	2	*+tA*	*+Vn*	*Vn*	*−n > −me*	1.6	✓
VIII	2	*+tA*	*+Vn*	*tOn*	*−n > −mA*	0.0	
IX	2	*+tA*	*+Vn*	*Us*	*−s > −kse*	4.7	✓
X	2	*+tA*	*+Vn*	*Ut**	*−t > −e*	0.0	
XI	2	*+tA*	*+seen*	*As**	*−s > −V*	1.6	✓
XII	2	*+ttA*	*+seen*	short *e*	*−e > −ee*	5.0	✓
XIII	3	*+tA*	*+Vn*	short *i*	*−i > −e*; *−i >* Ø	1.5	✓
XIV	3	*+tA*	*+Vn*	*nen*	*−nen > −se*; *−nen > −s*	2.6	✓
XV	3	*+tA*	*+Vn*	*si*	*−si > −te*; *−si > −t*	0.9	
XVI	3	*+tA*	*+Vn*	*[p,t,k]si*	*−i > −e*; *−Csi > −s*	0.2	
XVII	3	*+tA*	*+Vn*	*VVs*	*−s > −te*; *−s > −t*	0.8	

Table A.5
Summary of 12 general declension classes in Estonian, according to criteria used in the study (number of stems, genitive and partitive case formation, final phoneme in nominative form, type of stem changes) and the CDS corpus type frequency (%, including nouns and adjectives). 'V' (in the genitive and partitive columns) denotes a vowel which is lexically determined. All classes other than V, VII and IX-XII are considered productive. The final column indicates the classes included in the current study.

Class	Stems	Genitive	Partitive	Final	Stem changes	Type %	
I	1	Ø	+*d*	*vowel*	none	2.6	✓
II	2	+*(s)e*	+*(s)t*	−*e*/−*s*	if stem −*ne*, > −*se*	9.4	✓
III	1	Ø	+*t*	*vowel*	none	7.6	✓
IV	2	+*V*	+*Vt*	*consonant*	+ vowel	12.5	✓
V	1	Ø	Ø	*vowel*	none	9.5	✓
VI	2	weak stem + *V*	strong stem + *V*	*consonant*	weakening gradation	49.2	✓
VII	2	weak stem	Ø (strong stem)	*vowel*	weakening gradation	2.4	✓
VIII	2	strong stem + *V*	weak stem + *t*	*consonant*	strengthening gradation	2.8	✓
IX	2	+*me*	+*(n)t*	*vowel*	phonemic	1.0	
X	3	+*(n)e*	+*(n)t*	*consonant*	phonemic, +*e*	0.4	
XI	3	*i* > *e*	+*(n)d*	*vowel*	vowel change	0.5	
XII	3	weak stem, *i* > *e*	strong stem, *i* > *e*	*vowel*	weakening gradation, *i* > *e*	0.3	

declension classes are likely to be to some extent artificial and contain exceptions.

The Finnish Centre for Research in National Languages (*Kotimaisten kielten Tutkimuskeskus*, or *Kotus*) postulates 49 nominal declension classes based on nominal inflection patterns in both plural and singular contexts. However, as the current study investigates only singular contexts, a simplified classification system based on the Kotus criteria was devised, using only class differences relevant in singular contexts. This classification scheme led to the formation of 17 nominal classes (see Table A.4), divided according to four criteria: (a) number of stems per word; (b) the partitive affix required; (c) the illative affix required; and (d) any stem changes the word undergoes relative to the basic NOM stem. As can be seen in the table, the phonetic form of a nominal in the NOM case often reveals its declension class, although this is not always the case (For example, a nominal ending in a short –*I* may belong to class I, class V, class XIII, class XV or class XVI). Additionally, the phonological processes of consonant gradation and vowel harmony affect Finnish noun forms in the same way as they operate throughout the grammatical system.

The most authoritative dictionary of correct usage in Estonian, *Õigekeelsussõnaraamat* (Erelt, Leemets, Mäearu, & Raadik, 2018), distinguishes 26 main declension classes, with subclasses defined for differences in stem-final vowels used in case inflection. The genitive form always ends in a vowel, and that vowel-final form is used as the inflectional stem for 11 of 14 cases in the singular. For present purposes, we have used a simplified version of the classification system proposed by Kaalep (2012), based on the differences relevant to singular nouns; as in Finnish, distinctions only relevant in plural contexts were disregarded for this study. Twelve nominal declensions (see Table A.5) were classified based on: (a) final phoneme of the nominative form (vowel or consonant), (b) genitive formation, (c) partitive formation (+*d*, +*t*, or +Ø), and (d) type of stem change. The declension class is not always predictable from the phonological form of the nominative stem (see Blevins, 2008). Moreover, as can be seen in the table, some declension classes rely entirely on stem change rather than affixal endings to distinguish grammatical cases in Estonian, e.g. groups VI, VII, and XII in Table A.5.

Appendix B. Frequency distribution of case-marked nouns for each CDS corpus

See Fig. B.1.

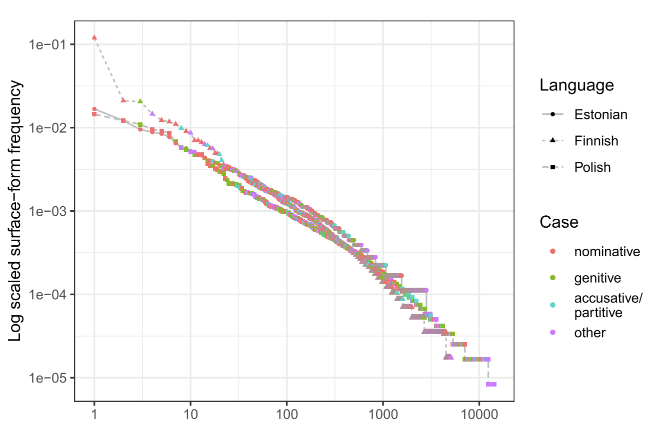


Fig. B.1. Log-log plot showing the relationship between frequency and rank of all noun surface forms in the CDS corpora.

Appendix C. Description and results of form-based PND measure (PND form)

C.1. Brief description of the neighbourhood calculation

Following suggestions from reviewers (especially Vera Kempe), we devised a continuous measure of PND which is based on the individual target form rather than the class membership of the noun (PND form). The new measure is represented by a ratio of summed similarities of neighbours vs. non-neighbours with respect to each individual target form. Neighbourhood is defined on the basis of similarity in terms of the *transformations* required to derive the particular target form (e.g., Polish *statk-u*, 'ship-GEN') from the nominative form of the same noun (e.g., *statek*). This process is similar in principle to the analogical model for the English past tense in Albright and Hayes (2003), which they based on the Generalized Context Model (Nakisa, Plunkett, & Hahn, 2001; Nosofsky, 1990). Our particular implementation is described briefly below.

For every noun form in the corpus, nominative → target form transformations are classified with regard to the number of phonemes inserted, deleted, affixed or substituted. Two NOM → target derivations are neighbours *if* (and only if) any shared deletions, insertions and substitutions differ by no more than one bit in their binary feature representation (explained below) AND they are identical with regard to (1) affixed phonemes, (2) ending on a vowel versus consonant and (3) the total number of transformations. Substitutions are calculated in terms of edit distance using a feature-based binary encoding scheme based on PatPho (Li & MacWhinney, 2002) For example, in Estonian, the change from /n/ to /s/ in /hobune/ → /hobuse/, ('horse', nominative → genitive), is a change of 3 bits in the feature representation (from voiced, alveolar nasal to unvoiced, alveolar fricative), while /p/ to /b/ in /se:p/ → /se:bi/ ('soap', nominative → genitive) is a change of 1 bit. The R package *alineR* (Downey, Sun, & Norquest, 2017) was used to find the optimal alignment between two forms by minimising the edit distance while also accounting for phoneme similarity. This makes sure that, for example, /pe:gel/ → /pe:gli/ ('mirror', nominative → genitive) is treated as a deletion of /e/ and affixation of /i/, rather than two substitutions, from /el/ to /li/.

Once neighbours of each target have been defined, a continuous similarity measure is calculated by summing across the similarity – defined in terms of edit distance and phoneme similarity using *alineR* – between the target and each of its neighbours. Finally, this summed similarity measure is normalized by dividing by the summed similarity between the relevant nominative form and the nominative forms of those nouns that are also present in the relevant case in the CDS corpus. As a result, the final PND measure is a ratio which is boosted by "friends" (nouns that are similar in both their nominative form and the required transformations to the target form; i.e., similar neighbours), but reduced by "enemies" (nouns that are similar in nominative form, but that differ in target transformations; i.e., similar non-neighbours).

C.2. Results of the analysis

See Table C.1.

Table C.1
Results of Bayesian analyses using the PND form measure. The estimated mean is reported as well as the lower and upper limits of the 95% credible interval and the probability *P* of the effect being < 0 (for negative coefficients) or > 0 (for positive coefficients) given the data. Boldface indicates strong evidence (Bayesian P ≥ 0.95).

	Mean	Lower	Upper	*P*
(a) Polish				
Intercept	1.96	1.57	2.37	1
PND form	**0.36**	**0.07**	**0.66**	**.99**
Surface-form frequency	**0.42**	**0.12**	**0.73**	**1**
Age	**0.57**	**0.27**	**0.86**	**1**
PND form:Surface-form frequency	**−0.36**	**−0.78**	**0.09**	**.95**
PND form:Age	**−0.02**	**−0.44**	**0.04**	**.95**
Surface-form frequency:Age	0.02	−0.24	0.28	.56
(b) Finnish				
Intercept	2.00	1.46	2.56	1
PND form	**0.32**	**0.04**	**0.63**	**.99**
Surface-form frequency	0.20	−0.08	0.47	.92
Age	**0.56**	**0.13**	**0.95**	**1**
PND form:Surface-form frequency	−0.10	−0.40	0.23	.74
PND form:Age	−0.18	−0.45	0.10	.91
Surface-form frequency:Age	0.10	−0.14	0.36	.78
(c) Estonian				
Intercept	3.90	2.94	4.91	1
PND form	**1.75**	**1.12**	**2.46**	**1**
Surface-form frequency	**0.33**	**−0.07**	**0.75**	**.95**
Age	**1.17**	**0.45**	**2.03**	**1**
PND form:Surface-form frequency	−0.16	−0.74	0.41	.71
PND form:Age	0.11	−0.58	0.79	.63
Surface-form frequency:Age	−0.12	−0.57	0.33	.71

Appendix D. Experimental stimuli

D.1. Criteria for experimental stimuli selection

Three nouns were selected from each declension class, as defined in Appendix A, to represent high, medium and low surface-form frequencies within the range of the class. We matched the selected nouns between languages as much as possible, in order to use similar pictorial stimuli across languages. As additional criteria, phonological properties (number of syllables and lack of consonant clusters) in Polish were taken into account where possible. In Finnish and Estonian, disyllabic noun stimuli were chosen as far as possible, and in Finnish, both front and back vowel contexts, as well as contexts with and without consonant gradation, were used. In Estonian, where applicable, nouns were selected with differing inflectional stem-final vowels and with varying phonological structure. Additionally, for Estonian, in noun class VI we excluded nouns in which the only cue to case distinctions is phoneme duration, as this renders it more difficult to assess production accuracy.

D.2. Nouns selected in each language

See Tables D.1–D.3.

Table D.1
Polish noun test items in nominative form.

Class	Nominative	Translation
Feminine1	*koszula*	'shirt'
	spódnica	'skirt'
	cebula	'onion'
Feminine2	*książka*	'book'
	szklanka	'glass'
	beczka	'barrel'
Masculine1	*pociąg*	'train'
	statek	'ship'
	dźwig	'crane'
Masculine2	*samochód*	'car'
	telefon	'phone'
	aparat	'camera'
Masculine3	*ołówek*	'pencil'
	garnek	'pot'
	słoik	'jar'
Masculine4	*chleb*	'bread'
	traktor	'tractor'
	grzebień	'comb'
Masculine5	*wilk*	'wolf'
	smok	'dragon'
	kucharz	'cook'
Masculine6	*doktor*	'doctor'
	bocian	'stork'
	wieloryb	'whale'
Neuter1	*jajko*	'egg'
	jabłko	'apple'
	sitko	'sieve'
Neuter2	*okno*	'window'
	koło	'wheel'
	pianino	'piano'

Table D.2
Finnish noun test items in nominative form.

Class	Nominative	Translation
I	*kirja*	'book'
	uuni	'oven'
	käpy	'pinecone'
III	*puu*	'tree'
	hai	'shark'
	vyö	'belt'
V	*ovi*	'door'
	putki	'pipe'
	kilpi	'shield'
VII	*puhelin*	'phone'
	avain	'key'
	kaulin	'rolling pin'
IX	*sormus*	'ring'
	sirkus	'circus'
	maalaus	'painting'

Table D.2 (*continued*)

Class	Nominative	Translation
XI	*allas*	'pool'
	rengas	'wheel'
	kiuas	'sauna stove'
XII	*kone*	'machine'
	hame	'skirt'
	aarre	'treasure'
XIII	*kuusi*	'spruce tree'
	sieni	'mushroom'
	kaari	'arch'
XIV	*lautanen*	'plate'
	hevonen	'horse'
	lapanen	'mitten'

Table D.3
Estonian noun test items in nominative form.

Class	Nominative	Translation
I	*kuu*	'moon'
	patarei	'battery'
	vöö	'belt'
II	*päike*	'sun'
	hobune	'horse'
	tsirkus	'circus'
III	*auto*	'car'
	voodi	'bed'
	raadio	'radio'
IV	*raamat*	'book'
	masin	'machine'
	madrats	'mattress'
V	*maja*	'house'
	toru	'pipe'
	kruvi	'screw'
VI	*täht*	'star'
	lamp	'lamp'
	sulg	'feather'
VII	*tuba*	'room'
	nuga	'knife'
	tigu	'snail'
VIII	*aken*	'window'
	ratas	'wheel'
	saabas	'boot'

D.3. Examples of picture stimuli

See Fig. D.1.

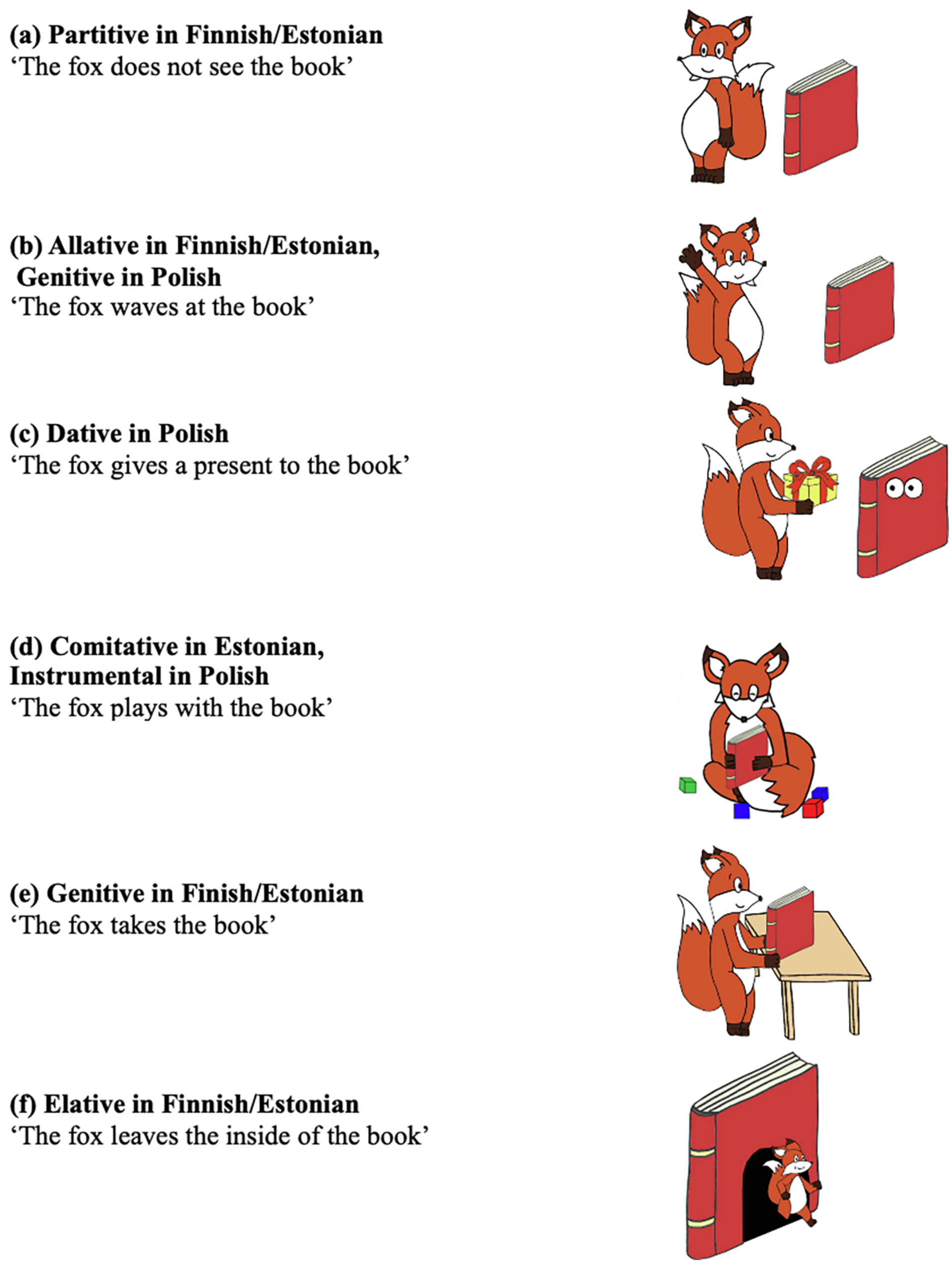


Fig. D.1. Examples of the elicitation contexts used in the study.

Appendix E. Exploratory analysis of language-specific factors

See Table E.1.

Table E.1
Results of the Bayesian exploratory analyses for the three languages. The estimated mean is reported as well as the lower and upper limits of the 95% credible interval and the probability *P* of the effect being < 0 (for negative coefficients) or > 0 (for positive coefficients) given the data. Boldface indicates strong evidence (Bayesian P ≥ 0.95).

	Mean	Lower	Upper	*P*
(a) Polish				
Intercept	2.16	1.65	2.68	1
PND class	1.50	−0.74	3.71	.91
Surface-form frequency	0.20	−0.13	0.52	.89
Syncretism	−0.04	−0.60	0.49	.55
Case variance	**−2.59**	**−3.35**	**−1.85**	**1**
PND class:Surface-form frequency	−1.11	−3.17	0.85	.86
PND class:Case variance	1.09	−2.07	4.56	.75
PND class:Syncretism	−0.78	−3.04	1.37	.75
Surface-form frequency:Case variance	0.05	−1.06	1.14	.54
Surface-form frequency:Syncretism	0.26	−0.31	0.82	.83
Syncretism:Case variance	**−2.35**	**−5.05**	**0.17**	**.97**
(b) Finnish				
Intercept	**3.88**	**3.02**	**4.81**	**1**
PND class	**2.99**	**1.22**	**4.88**	**1**
Surface-form frequency	0.16	−0.22	0.52	0.8
Stem-change	−0.2	−0.84	0.42	0.73
Case variance	**−1.14**	**−1.64**	**−0.64**	**1**
PND class:Surface-form frequency	0.08	−0.31	0.47	0.65
PND class:Stem-change	**3.29**	**1.54**	**5.16**	**1**
PND class:Case variance	**0.45**	**−0.09**	**1.02**	**0.95**
Surface-form frequency:Stem-change	0.15	−0.28	0.59	0.76
Surface-form frequency:Case variance	**−0.36**	**−0.76**	**0.03**	**0.97**
Stem-change:Case variance	**0.51**	**0.04**	**0.99**	**0.98**
(c) Estonian				
Intercept	4.7	3.78	5.7	1
PND class	**5.09**	**2.38**	**7.64**	**1**
Surface-form frequency	**0.56**	**0.06**	**1.08**	**0.99**
Stem-change	**−1.25**	**−1.76**	**−0.76**	**1**
Case variance	−0.46	−1.11	0.21	0.91
PND class:Surface-form frequency	0.7	−1.2	2.53	0.78
PND class:Stem-change	0.83	−1.07	2.7	0.8
PND class:Case variance	**−2.68**	**−5.1**	**−0.39**	**0.99**
Surface-form frequency:Stem-change	−0.35	−1	0.22	0.87
Surface-form frequency:Case variance	−0.69	−1.55	0.14	0.94
Stem-change:Case variance	**1.55**	**0.9**	**2.35**	**1**

Appendix F. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2019.04.004.

References

Aguado-Orea, J., & Pine, J. M. (2015). Comparing different models of the development of verb inflection in early child Spanish. *PLoS One, 10*, e0119613.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition, 90*, 119–161.

Alegre, M., & Gordon, P. (1999). Rule-based versus associative process in derivational morphology. *Brain and Language, 68*, 347–354.

Argus, R. (1998). CHILDES'i eesti andmepank ja selle suhtluskeskne analüüs (Hendrik, 1.6-2.6) ['Overview and analysis of the Estonian databank on CHILDES, (Hendrik, 1;6-2;6)']. (Unpublished Master's thesis). Tallinn, Estonia: Tallinn Pedagogical University.

Argus, R. (2009b). The early development of case and number in Estonian. In M. D. Voeikova, & U. Stephany (Eds.). *Development of nominal inflection in first language acquisition. A cross-linguistic perspective* (pp. 111–152). Berlin: Mouton De Gruyter.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Baayen, R. H., McQueen, J., Dijkstra, T., & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In R. H. Baayen, & R. Schreuder (Eds.). *Morphological structure in language processing* (pp. 355–390). Berlin, Germany: Mouton de Gruyter.

Baayen, R. H., Wurm, L. H., & Aycock, J. (2007). Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. *The Mental Lexicon, 2*, 419–463.

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language, 36*, 94–117.

Baayen, R. H., Kuperman, V., & Bertram, R. (2010). Frequency effects in compound processing. In S. Scalise, & I. Vogel (Eds.). *Compounding* (pp. 257–270). Amsterdam/Philadelphia: Benjamins.

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01.

Baus, C., Costa, A., & Carreiras, M. (2008). Neighbourhood density and frequency effects in speech production: A case for interactivity. *Language and Cognitive Processes, 23*, 866–888.

Bittner, D., Dressler, W. U., & Kilani-Schoch, M. (Vol. Eds.), (2003). *Development of verb inflection in first language acquisition: A cross-linguistic perspective: Vol. 21*. New York: Walter de Gruyter.

Blevins, J. P. (2008). Declension classes in Estonian. *Linguistica Uralica, XLIV*, 241–267.

Bürkner, P. C. (2018). Advanced Bayesian multilevel modelling with the R Package brms.

The R Journal, 10, 395–411.
Bybee, J. L., & Slobin, D. I. (1982). Rules and schemas in the development and use of the English past tense. *Language, 58*, 265–289.
Cameron-Faulkner, T., & Kidd, E. (2007). I'm are what I'm are: The acquisition of first-person singular present BE. *Cognitive Linguistics, 18*, 1–22.
Chandler, S. (2010). The English past tense: Analogy redux. *Cognitive Linguistics, 21*, 371–417.
Clahsen, H., Rothweiler, M., Woest, A., & Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition, 45*, 225–255.
Dąbrowska, E. (2004). Rules or schemas? Evidence from Polish. *Language and Cognitive Processes, 19*, 225–271.
Dąbrowska, E. (2008). The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language, 58*, 931–951.
Dąbrowska, E., & Szczerbiński, M. (2006). Polish children's productivity with case marking: The role of regularity, type frequency, and phonological diversity. *Journal of Child Language, 33*, 559–597.
Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems, 15*, 1150017.
Deen, K., & Hyams, N. (2006). The morphosyntax of mood in early grammar with special reference to Swahili. *First Language, 26*, 67–102.
Downey, S. S., Sun, G., & Norquest, P. (2017). alineR: an R Package for optimizing feature-weighted alignments and linguistic distances. *The R Journal, 9*, 138–152.
Dressler, W. U. (2003). Degrees of grammatical productivity in inflectional morphology. *Italian Journal of Linguistics, 15*, 31–62.
Elman, J. L., Bates, E. A., & Johnson, M. H. (1998). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.
Engelmann, F., Granlund, S., Kolak, J., Szreder, M., Ambridge, B., Pine, J., ... Lieven, E. (2019). How the input shapes the acquisition of verb morphology: Elicited production and computational modelling in two highly inflected languages. *Cognitive Psychology, 110*, 30–69.
Engelmann, F., Granlund, S., Kolak, J., Vihman, V., Ambridge, B., Pine, J., … & Lieven, E. (2018). How the input shapes the acquisition of inflectional morphology: Modelling across three highly inflected languages. In Abstract book of the 43rd annual Boston University Conference on Language Development, Boston, MA.
Erelt, M., Erelt, T., & Ross, K. (2000). *Eesti keele käsiraamat [Handbook of Estonian].* Tallinn: Eesti keele sihtasutus.
Erelt, T., Leemets, T., Mäearu, S., & Raadik, M. (2018). *Eesti õigekeelsussõnaraamat [Dictionary of Correct Usage in Estonian].* Tallin: Eesti keele sihtasutus Retrieved from http://www.eki.ee/dict/qs/.
Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician, 60*, 328–331.
Gruszczyński, W., & Bralczyk, J. (2002). *Słownik gramatyki języka polskiego [Grammatical Dictionary of Polish].* Warszawa: WSIP.
Haman, E., Etenkowski, B., Łuniewska M., Szwabe, J., Dąbrowska, E., Szreder, M., & Łaziński, M. (2011). The Polish CDS Corpus. Retrieved from http://childes.psy.cmu.edu/data/Slavic/Polish/CDS.Zip.
Hartshorne, J. K., & Ullman, M. T. (2006). Why girls say 'holded'more than boys. *Developmental Science, 9*(1), 21–32.
Hoekstra, T., & Hyams, N. (1998). Aspects of root infinitives. *Lingua, 106*, 81–112.
Janssen, B. (2016). *The acquisition of gender and case in Polish and Russian: A study of monolingual and bilingual children.* Amsterdam: Uitgeverij Pegasus.
Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29*, 687–715.
Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences United States of America, 96*, 7592–7597.
Kaalep, H.-J. (2012). Käänamissüsteemi seaduspärasused [Regularities in the Estonian declension system]. *Keel ja Kirjandus, 6*, 418–449.
Kempe, V., Brooks, P. J., Mironova, N., Pershukova, A., & Fedorova, O. (2007). Cross-linguistic evidence for the diminutive advantage: Gender agreement in Russian and Serbian children. *Journal of child language, 34*, 111–131.
Kirjavainen, M., Kidd, E., & Lieven, E. (2017). How do language-specific characteristics affect the acquisition of different relative clause types? Evidence from Finnish. *Journal of Child Language, 44*, 120–157.
Kirjavainen, M., Nikolaev, A., & Kidd, E. (2012). The effect of frequency and phonological neighbourhood density on the acquisition of past tense verbs by Finnish children. *Cognitive Linguistics, 23*, 273–315.
Kjærbæk, L., dePont Christensen, R., & Basbøll, H. (2014). Sound structure and input frequency impact on noun plural acquisition: Hypotheses tested on Danish children across different data types. *Nordic Journal of Linguistics, 37*, 47–86.
Krajewski, G., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2011). How Polish children switch from one case to another when using novel nouns: Challenges for models of inflectional morphology. *Language and Cognitive Processes, 26*, 830–861.
Laaha, S., & Dressler, W. U. (2012). Suffix predictability and stem transparency in the acquisition of German noun plurals. In F. Kiefer, M. Ladanyi, & P. Siptar (Eds.). *Current issues in morphological theory:(ir)regularity, analogy and frequency* (pp. 217–235). Amsterdam: Benjamins.
Laaha, S., & Gillis, S. (2007). Typological perspectives on the acquisition of noun and verb morphology. In S. Laaha, & S. Gillis (Eds.). *Antwerp papers in linguistics 112*. Antwerp: Antwerp University.
Laalo, K. (2001). Diminutives in Finnish child-directed and child speech: Morphopragmatic and morphophonemic aspects. *Psychology of Language and Communication, 5*, 71–80.
Laalo, K. (2009). Acquisition of case and plural in Finnish. In M. D. Voeikova, & U. Stephany (Eds.). *Development of nominal inflection in first language acquisition: A cross-linguistic perspective* (pp. 87–182). Berlin: Mouton De Gruyter.
Leonard, L. B., Caselli, M. C., & Devescovi, A. (2002). Italian children's use of verb and noun morphology during the preschool years. *First Language, 22*, 287–304.
Li, P., & MacWhinney, B. (2002). PatPho: A phonological pattern generator for neural networks. *Behavioural Research Methods, Instruments, & Computers, 34*, 408–415.
Lõo, K., Järvikivi, J., & Baayen, R. H. (2018). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition, 175*, 20–25.
Lõo, K., Järvikivi, J., Tomaschek, F., Tucker, B. V., & Baayen, R. H. (2018). Production of Estonian case-inflected nouns shows whole-word frequency and paradigmatic effects. *Morphology, 28*, 71–97.
Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PloS one, 5*, e8559.
Marchman, V. A., Plunkett, K., & Goodman, J. (1997). Overregularization in English plural and past tense inflectional morphology: A response to Marcus (1995). *Journal of Child Language, 24*, 767–779.
Marchman, V. A., Wulfeck, B., & Weismer, S. E. (1999). Morphological productivity in children with normal language and SLI: A study of the English past tense. *Journal of Speech, Language, and Hearing Research, 42*, 206–219.
Maslen, R. J. C., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2004). A dense corpus study of past tense and plural overregularization in English. *Journal of Speech, Language and Hearing Research, 47*, 1319–1333.
Matthews, D., & Theakston, A. L. (2006). Errors of omission in English speaking children's production of plurals and the past tense: The effects of frequency, phonology, and competition. *Cognitive Science, 30*, 1027–1052.
McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences, 6*, 465–472.
Meunier, F., & Marslen-Wilson, W. D. (2004). Regularity and irregularity in French verb inflection. *Language and Cognitive Processes, 19*, 561–580.
Mirković, J., Seidenberg, M. S., & Joanisse, M. F. (2011). Rules versus statistics: Insights from a highly inflected language. *Cognitive Science, 35*, 638–681.
Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review, 23*, 103–123.
Nakisa, R. C., Plunkett, K., & Hahn, U. (2001). A cross-linguistic comparison of single and dual-route models of inflectional morphology. In P. Broeder, & J. Murre (Eds.). *Models of language acquisition: Inductive and deductive approaches*. Cambridge, MA: MIT Press.
Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—part II. *Language and Linguistics Compass, 10*, 591–613.
Nosofsky, R. M. (1990). Relations between exemplar similarity and likelihood models of classification. *Journal of Mathematical Psychology, 34*, 393–418.
Orsolini, M., & Marslen-Wilson, W. (1997). Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes, 12*, 1–47.
Pham, H., & Baayen, R. H. (2015). Vietnamese compounds show an anti-frequency effect in visual lexical decision. *Language, Cognition and Neuroscience, 30*, 1077–1095.
Pinker, S. (1999). *Words and rules: The Ingredients of Language.* New York: Basic Books.
Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in Cognitive Science, 1*(6), 456–463.
Plunkett, K., & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science, 23*, 463–490.
Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition, 38*, 43–102.
Plunkett, K., & Marchman, V. (1993). From rote learning to system-building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48*, 21–69.
Pothos, E. M. (2005). The rules versus similarity distinction. *Behavioral and Brain Sciences, 28*, 1–14.
Prasada, S., & Pinker, S. (1993). Generalizations of regular and irregular morphology. *Language and Cognitive Processes, 8*, 1–56.
R Core Team (2016). *R: A language and environment for statistical computing.* Austria: R Foundation for Statistical Computing Vienna Retrieved from https://www.R-project.org/.
Räsänen, S. H. M., Ambridge, B., & Pine, J. M. (2014). Infinitives or bare stems? Are English-speaking children defaulting to the highest-frequency form? *Journal of Child Language, 41*, 756–779.
Räsänen, S. H. M., Ambridge, B., & Pine, J. M. (2016). An elicited production study of inflectional verb morphology in child Finnish. *Cognitive Science, 40*, 1704–1738.
Sassenhagen, J., & Alday, P. M. (2016). A common misapplication of statistical inference: Nuisance control with null-hypothesis significance tests. *Brain and Language, 162*, 42–45.
Saviciute, E., Ambridge, B., & Pine, J. M. (2018). The roles of word-form frequency and phonological neighbourhood density in the acquisition of Lithuanian noun morphology. *Journal of Child Language, 45*, 641–672.
Savickienė, I., Kempe, V., & Brooks, P. J. (2009). Acquisition of gender agreement in Lithuanian: Exploring the effect of diminutive usage in an elicited production task. *Journal of Child Language, 36*, 477–494.
Schuler, K. D., Yang, C., & Newport, E. L. (2016). Testing the tolerance principle: Children form productive rules when it is more computationally efficient to do so. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.). *Proceedings of the 38th annual conference of the cognitive science society*. Austin, TX: Cognitive Science Society.
Seidenberg, M. S., & Gonnerman, L. M. (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Science, 4*, 353–361.
Slobin, D. I. (1985). *The crosslinguistic study of language acquisition, vol.1: The data.* Hillsdale, N.J.: Lawrence Erlbaum.
Sorensen, T., Hohenstein, S., & Vasishth, S. (2016). Bayesian linear mixed models using

Stan: A tutorial for psychologists, linguists, and cognitive scientists. *The Quantitative Methods for Psychology, 12*, 175–200.
Stan Development Team (2018). Stan Modeling Language Users Guide and Reference Manual, Version 2.18.0. http://mc-stan.org.
Stephany, U., & Voejkova, M. D. (2009). Development of nominal inflection in first language acquisition: A cross-linguistic perspective. In U. Stephany, & M. D. Voejkova (Eds.). *Studies on Language Acquisition 30*. Berlin: Mouton de Gruyter.
Stoll, S., Mazara, J., & Bickel, B. (2017). The acquisition of polysynthetic verb forms in Chintang. In M. D. Fortescue, M. Mithun, & N. Evans (Eds.). *The Oxford handbook of polysynthesis* (pp. 495–514). Oxford: Oxford University Press.
Theakston, A. L., & Lieven, E. V. M. (2005). The acquisition of auxiliaries BE and HAVE: An elicitation study. *Journal of Child Language, 32*, 587–616.
Theakston, A. L., & Rowland, C. F. (2009). The acquisition of auxiliary syntax: A longitudinal elicitation study. Part I: Auxiliary BE. *Journal of Speech, Language, and Hearing Research, 46*, 1449–1470.
Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2005). The acquisition of auxiliary syntax: BE and HAVE. *Cognitive Linguistics, 16*, 247–277.
Thymé, A., Ackerman, F., & Elman, J. (1994). Finnish nominal inflection. Paradigmatic patterns and token analogy. In S. D. Lima, R. L. Corrigan, & G. K. Iverson (Eds.). *The reality of linguistic rules* (pp. 445–466). Amsterdam: John Benjamins.
Tokarski, J. (1973). *Fleksja polska [Polish inflection]*. Warszawa: Polskie Wydawnictwo Naukowe.
Vija, M. (2007). Pronoomenid lapsekeeles: mõnda MINA ja SINA omandamisest eesti laste näitel. [Pronouns in child language: Some thoughts on the acquisition of I and YOU, based on Estonian]. *Estonian Papers in Applied Linguistics, 3*, 373–384.
Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 735–747.
Vitevitch, M. S., & Sommers, M. S. (2003). The facilitative influence of phonological similarity and neighborhood frequency in speech production in younger and older adults. *Memory & Cognition, 31*, 491–504.
Vitevitch, M. S., & Luce, P. A. (2016). Phonological neighborhood effects in spoken word perception and production. *Annual Review of Linguistics, 2*, 7.1–7.20.
Vitevitch, M. S., & Stamer, M. K. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes, 21*, 760–770.
Westermann, G., & Ruh, N. (2012). A neuroconstructivist model of past tense development and processing. *Psychological Review, 119*, 649–667.
Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua, 106*, 23–79.
Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koc, A., Christofidou, A., ... Dressler, W. U. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Language, 31*, 461–479.
Zupping, S. (2016). Zupping Corpus available on CHILDES. https://doi.org/10.21415/T5K89H.