A method for modelling nominal defectiveness in an Icelandic corpus

Alexandra Lilia Ayala
University of British Columbia

Abstract: A word, or class of words, is said to be defective if native speakers are reluctant to produce an inflected form. Oftentimes, different lexical items may be missing pronounced forms for the same feature or combination of features. This is the sort of systematic defectiveness that emerges as a morphological gap, and crucially depends on a native speaker's expectation that a pronounced form ought to exist. Under generative linguistic approaches, speakers cannot learn from negative evidence, and so these expectations need to be inferred through analogy, such as non-gapped lemmas with phonologically similar stem shapes. This is a relational puzzle whereby form-meaning mapping fails for some lexical items but not others. Native speakers are often not aware that a gap exists until it is encountered in a context that licenses its intended form. This poses a problem for gap discovery and identification, and it is a barrier to our understanding of where and why defectiveness happens. The first objective of this qualifying paper is to prod the relationship between gaps and low token frequencies by observing known gapped Icelandic lemmas. This study employs frequentist statistical models and finds empirical support for two causes of defectiveness in the genitive plural in the weak feminine inflection class: uncertainty over suffix choice and syncretism avoidance between the genitive plural and nominative singular. The second objective of this paper is to tease out the predictors that correlate with low token frequencies. The factors identified in this qualifying paper (inflection class, case, definiteness, and lemma frequency) can be generalised in future work as a tool for diagnosing defectiveness.

1 Introduction

Icelandic speaker-linguists have long expressed intuitions that there are GEN.PL word forms missing from a particular subset of the Icelandic lexicon (Cleasby and Vigfusson 1874; Hansson 2015; Jónsson 1927). In this qualifying paper, I will show how morpho-phonological divisions in Icelandic inflectional categories have conspired to induce instability in the GENITIVE PLURAL of a particular inflectional class. In this section, I will lay out the ingredients for what will eventually lead to these pockets of defectiveness, starting with the inflectional classes themselves. Icelandic belongs to the North Germanic language family and is mostly spoken in Iceland by approximately 400,000 speakers (Eberhard et al. 2021). It is a synthetic language with rich inflectional morphology expressed primarily through suffixation. Nouns are inherently gendered as either feminine, masculine, or neuter. Noun suffixes inflect for case (nominative, accusative, dative, genitive) and number (singular, plural). Definiteness is marked with an enclitic determiner that affixes to the rightmost edge of the inflected stem. The determiner also inflects for case and number, as well as grammatical gender.

Within each gender, there are multiple inflection classes whose morphosyntactic and phonological properties overlap in various ways (Ingason 2016). Table 1 exhibits one theory of how these inflection classes can be carved up, with nine strong classes and three weak classes. These classes are oftentimes referred to by a thematic vowel (e.g., "Ma" in table 1 refers to the strong masculine class of nouns that are analysed by Müller (2005) as being defined by the *a* thematic vowel). In each cell, the exponences for each indefinite case/number combination are shown. The defective inflection class that will be examined in this paper can be seen in the last emboldened column of table 1, and is called the "weak feminine" (henceforth Fw) inflection class[1].

Table 1: Inflection class suffixes in Icelandic nouns (Müller 2005:235)

	Strong									Weak		
	Ma	Na	Fa(′)	Mi	Fi	Mu	Mc	Fc1	Fc2	Mw	Nw	**Fw**
NOM.SG	-ur	∅	∅	-ur	∅	-ur	-ur	∅	∅	-i	-a	**-a**
ACC.SG	∅	∅	∅(-u)	∅	∅	∅	∅	∅	∅	-a	-a	**-u**
DAT.SG	-i	-i	∅(-u)	∅	∅	-i	-i	∅	∅	-a	-a	**-u**
GEN.SG	-s	-s	-ar	-ar	-ar	-ar	-ar	-ar	-ur	-a	-a	**-u**
NOM.PL	-ar	∅	-ar	-ir	-ir	-ir	-ur	-ur	-ur	-ar	-u	**-ur**
ACC.PL	-a	∅	-ar	-i	-ir	-i	-ur	-ur	-ur	-a	-u	**-ur**
DAT.PL	-um	-um	-um	-um	-um	-um	-um	-um	-um	-um	-um	**-um**
GEN.PL	-a	-a	-a	-a	-a	-a	-a	-a	-a	-a	-(n)a	**-(n)a**

The Fw and Mw are both large, open, and productive inflection classes. They both exhibit a high degree of syncretism and they can both form the GEN.PL with the suffix *-a*, as do nearly all other classes. Many neologisms and borrowings can often be found in these classes. Nouns that belong to the weak neuter (Nw) class are very few in number and have a more restricted semantic distribution; these are often organs and other body parts. Like the Fw, the GEN.PL can be formed with either suffix *-na* or *-a*. Table 2 below lists the indefinite and definite inflected forms of three lexical items (represented by the lemmas BENDA, ENDI, and LUNGA) in each class.

Table 2: Sample of weak nominal paradigms

	Fw		Mw		Nw	
	BENDA 'tangle'		ENDI 'end'		LUNGA 'lung'	
	INDF	DEF	INDF	DEF	INDF	DEF
NOM.SG	bend-a	bend-a-n	end-i	end-i-nn	lung-a	lung-a-ð
ACC.SG	bend-u	bend-u-na	end-a	end-a-nn	lung-a	lung-a-ð
DAT.SG	bend-u	bend-u-nni	end-a	end-a-num	lung-a	lung-a-nu
GEN.SG	bend-u	bend-u-nnar	end-a	end-a-ns	lung-a	lung-a-ns
NOM.PL	bend-ur	bend-ur-nar	end-ar	end-ar-nir	lung-u	lung-u-n
ACC.PL	bend-ur	bend-ur-nar	end-a	end-a-na	lung-u	lung-u-n
DAT.PL	bend-um	bend-u-num	end-um	end-u-num	lung-um	lung-u-num
GEN.PL	bend-(n)a	bend-(n)a-nna	end-a	end-a-nna	lung-na	lung-na-nna

The two suffixes *-a* and *-na* that are available to Fw nouns are distributed in a semi-predictable way. There is a set of stem shapes that are characterised by near-categorical preference for taking *-a* with near-certainty. This includes stems ending in any of the following: a geminate or singleton consonant + /r/, a singleton /n/, a non-dorsal consonant followed by /j/, and a geminate /n/ or a consonant + /n/ cluster. From here on out, this phonologically-defined set will be referred to as the N-FREE stem type, whereas all other stem shapes will be referred to as OTHER. These other stems

[1] In this table and throughout this paper, orthographic representations are used. When relevant, phonological representations will be given.

that belong to the OTHER set can take *-a*, *-na*, or both *-(n)a*, but it is not predictable from stem shape alone which suffix such stems would select.

Icelandic speakers demonstrate a general reluctance or hesitance toward producing the GEN.PL forms of some Fw nouns. There is disagreement over which lemmas are missing GEN.PL forms, and there is no sufficiently generalisable phonological or morphological pattern that can be used to predict which lexical items will be affected. When a lemma in a language is missing an inflected form that can otherwise be expressed in other paradigms, the phenomenon is known as defectiveness. One example of a defective paradigm is shown below in Table 3, where the defective forms (forms which are expected but unacceptable to many speakers) are marked with '?'.

Table 3: Defective paradigm for ALDA 'wave'

ALDA	INDF.SG	DEF.SG	INDF.PL	DEF.PL
NOM	ald-a	ald-a-n	öld-ur	öld-ur-nar
ACC	öld-u	öld-u-na	öld-ur	öld-ur-nar
DAT	öld-u	öld-u-nni	öld-um	öld-u-num
GEN	öld-u	öld-u-nnar	**?ald-a / ?ald-na**	ald-a-nna / **?ald-na-nna**

The defectiveness that Cleasby and Vigfusson (1874) reported for this particular lemma was restricted to the *-a* and *-na* forms of the INDF.GEN.PL. In their Icelandic-English dictionary, they write, "Only a few of the words of this declension (little more than a score, or about two or three per cent. of the whole) form a gen. plur." (Cleasby and Vigfusson 1874:18). The authors go on to say that other nouns of this inflection class form the GEN.PL by borrowing their inflected forms from the NOM.SG in single stem words, or the GEN.SG in compounds. Some cases like ALDA, which the authors say lack a "proper" GEN.PL form, are said to be possible in the definite. It should be noted that these judgements may vary across texts and speakers, and although such reports cannot be directly interpreted as the absence of certain forms, they still suggest the presence of defectiveness.

Since diagnosing defectiveness requires negative evidence, it can be very challenging, if not impossible, to elicit evidence of defectiveness directly from native speakers. Many are often not even aware that a lemma is defective until a context is encountered that requires an expected but unavailable form. For instance, while the sentence in (1a) would be perfectly acceptable, a speaker would hesitate to produce (1b), and would likely even opt to change the sentence[2].

(1) a. Ég hitti eiganda þessara blaðra.
je:g hɪht-ɪ eɪ:ɣant-a θes:ar-a blaðr-a
1.SG.NOM meet-1.SG.PST owner-ACC.SG that-GEN.PL balloon-GEN.PL
'I met the owner of those balloons.' (GÓH)

b. Ég hitti eiganda þessara **?fiðla**.
je:g hɪht-ɪ eɪ:ɣant-a θes:ar-a fɪðl-a
1.SG.NOM meet-1.SG.PST owner-ACC.SG that-GEN.PL violin-GEN.PL
'I met the owner of those violins.' (GÓH)

There are two puzzles that arise from these data: how to generalise the factors that correlate with defectiveness in Icelandic and how to measurably quantify defectiveness. Statistical models based

[2] Judgements given by Gunnar Ólafur Hansson (henceforth GÓH).

on corpus data are an ideal avenue to test these puzzles. First, the intuitions of Icelandic speaker-linguists can be used to form hypotheses about which phonological or morphological factors correlate most strongly with defectiveness, and defectiveness itself can be defined as either gradient or categorical statistical underrepresentation. Second, corpora have both a large amount of language data as well as a somewhat naturalistic distribution of data, more so at least than prescriptive sources like dictionaries and handbooks. Corpora are also advantageous because unlike in elicitation contexts, avoidance can be identified as trends across many different speakers, registers, time periods, and genres. This is somewhat a double-edged sword: while this variation adds layers of complexity to the generalisability of results, it also makes a stronger case for defectiveness if there are trends of non- or under-attestation despite such diversity in sources.

2 Theoretical frameworks and data

2.1 Morphological assumptions

For convenience, this study assumes that paradigms are atomic linguistic objects. This assumption comes from the morphological theory known as Word-and-Paradigm, which posits a *part-whole relation* where the smallest meaningful unit is the word (Blevins 2016; Blevins et al. 2018; Stump 2001). Crucially, knowledge of one word in a paradigm can be used to infer other words from the same paradigm (Ackerman et al. 2009; Finkel and Stump 2007). This means that certain words may be more informative than others. Highly irregular forms and suppleted forms, for example, would be much less informative than more regularly inflected words.

Lemmas, often written in capital letters, are abstractions of each paradigm, much like a set name variable abstracts a set. The parts of each set are organised by their *morphosyntactic* and *morphotactic* properties. Morphosyntactic properties are determined by features, while morphotactic properties can be thought of as the realisations of those features. While an entire word may be the morphosyntactic atomic unit, the morphotactic unit may be smaller, like a stem or a formative, or larger, like a periphrastic expression. Unlike in constructionist models of morphology (e.g., Distributed morphology; Halle and Marantz (1993)), which assume one-to-one correspondences between meaning and form, Word-and-Paradigm models make no such assumption; for instance, in Icelandic, case and number are two features (or feature bundles) that define a morphosyntactic unit, but they are only realised in a single morphotactic unit. This asymmetry is the difference between a null realisation and defectiveness, where a morphosyntactic unit (i.e., a cell) has no morphotactic expression at all (i.e., an empty cell).

Interestingly, native speakers generally know what form a word would take if its cell were not empty. This knowledge can either come from relational facts inferred from other parts of the same paradigm, or by analogy from relations within other paradigms. Nonetheless, relational awareness is important not only for determining how to fill a cell, but also for establishing when cells are empty. For instance, there is no locative case in Icelandic. Therefore, the lack of LOC wordforms is not attributable to a gap in the language, since there are no available morphosyntactic features to express the LOC anyway. Conversely, there *is* a GEN case, and so by necessity there are many affixes that express the morphosyntactic features of the GEN. This is why native speakers can perceive that a cell is "empty," rather than a word that simply does not belong. In count data, it is difficult to tease apart the distinction between unfamiliar words, which are unexpected, and defective words, which are expected, because both would manifest infrequently. Part of the goal of this methodology is to determine this difference.

2.2 Possible causes of defectiveness in Icelandic Fw nouns

Traditional descriptions of Icelandic have attributed Fw GEN.PL defectiveness mainly to two converging factors: syncretism avoidance and indeterminacy. I will be exploring these two potential causes in-depth, though in this section I will further make mention of phonotactic ill-formedness and aversion to innovative allomorphy, both of which are of particular concern to stems that take *-na* in the GEN.PL. Since the results presented in section 5 do not consider phonological factors, section 6 will show how different levels of word formation can interact and cause different kinds of defectiveness in the same paradigm cell. The interaction of these many ingredients lead to problematic inflection strategies for which I will argue at the end of this paper that there is no existing mechanism that can adequately capture or predict. For each of these possible causes, I will present some data from Icelandic, and then I will map those data to existing theoretical frameworks in defectiveness literature.

2.2.1 Syncretism avoidance

The first potential factor is homophony avoidance: Fw nouns that take *-a* in the genitive plural are homophonous with the nominative singular. This means that lexical items that have the N-FREE stem type will categorically exhibit homophony between the NOM.SG and GEN.PL. For example, in table 4 below, the GEN.PL form of *stjarna* is identical to its NOM.SG form. This identity relation is no longer a problem when the definite marker is encliticised. Notably, this is not the only case of homophony in table 4 below, but as I will discuss in the next paragraph as well as later on in section 6, the NOM.SG and GEN.PL never overlap in their syntactic distributions. However, the NOM.SG is highly frequent and plays a central role, both as the default form for syntactic subjects and a citation form for Icelandic speakers.

Table 4: Defective paradigm for STJARNA 'star'

STJARNA	SG.INDF	SG.DEF	PL.INDF	PL.DEF
NOM	**stjarn-a**	stjarn-a-n	stjörn-ur	stjörn-ur-nar
ACC	stjörn-u	stjörn-u-na	stjörn-ur	stjörn-ur-nar
DAT	stjörn-u	stjörn-u-nni	stjörn-um	stjörn-u-num
GEN	stjörn-u	stjörn-u-nnar	**?stjarn-a**	stjarn-a-nna

Müller (2005) and Baerman (2011) both point out that the syncretism between the NOM.SG and the GEN.PL is a privileged correspondence that could result in defectiveness of the less prominent cell because of a violation of the *syncretism principle*. This principle states that a single form should map to a single meaning, and so once such a relation is established, a language learner would ostensibly avoid mapping another meaning to the learned form. This particular pattern in syncretism in Icelandic furthermore has the property of collapsing two disjoint sets of features, namely, in both case and number. Elsewhere in the Fw paradigm, there is syncretism within the same number (e.g., ACC.SG, DAT.SG, and GEN.SG). In other classes, there is also syncretism within the same case (e.g., Mw ACC.SG and ACC.PL), but identity between the Fw NOM.SG and GEN.PL is a relation between cells that share neither case nor number.

2.2.2 Indeterminacy

The second factor is indeterminacy. The OTHER stem type does not entail a particular suffix choice in the GEN.PL. If an Icelandic speaker encounters one of these OTHER stems, then there is no low-level strategy that a speaker can use to be sure about which suffixal allomorph that stem can take; this information must be provided elsewhere. In sentence (2), *sveðja* belongs to this stem type that unambiguously takes *-a*, which will henceforth be called the N-FREE stem type.

(2)

Þeir	bönnuðu	sölu	sveðja.
θei:r	pœn:-YðY	sœ:l-Y	sveðj-a
3.M.NOM.PL	ban-3.PL.PST	sale-ACC.SG	machete-GEN.PL

‘They banned the sale of machetes.’ (Translation by GÓH)

Some examples of the OTHER stem type are shown in sentences like (3a). The noun in sentence (3a) takes *-na* in the GEN.PL, but this suffix is not predictable from the shape of the stem *kúl-*. Some OTHER stems can exhibit optionality, where the same stem can be suffixed either by *-a* or *-na*, as in (3b). This optionality can occur in free variation in some speakers' productions, while other speakers may prefer to use one variant more than the other.

(3) a.

Þeir	bönnuðu	sölu	kúlna.
θei:r	pœn:-YðY	sœ:l-Y	kul-na
3.M.NOM.PL	ban-3.PL.PST	sale-ACC.SG	bullet-GEN.PL

‘They banned the sale of bullets.’ (Translation by GÓH)

b.

Þeir	bönnuðu	sölu	byssna/byssa.
θei:r	pœn:-YðY	sœ:l-Y	pɪs-(n)a
3.M.NOM.PL	ban-3.PL.PST	sale-ACC.SG	gun-GEN.PL

‘They banned the sale of guns.’ (Translation by GÓH)

Some nouns of the OTHER stem type contain stem-final clusters that result in an ill-formed sequence of consonants if *-na* is selected as the suffix in the GEN.PL. For instance, in table 5, the *-na* option would result in a cluster [ŋln]. However, selecting the alternative suffix *-a* would result in homophony with the NOM.SG.

Table 5: Defective paradigm for KRINGLA ‘disk; pretzel; discus’

KRINGLA	SG.INDF	SG.DEF	PL.INDF	PL.DEF
NOM	**kringl-a**	kringl-a-n	kringl-ur	kringl-ur-nar
ACC	kringl-u	kringl-u-na	kringl-ur	kringl-ur-nar
DAT	kringl-u	kringl-u-nni	kringl-um	kringl-u-num
GEN	kringl-u	kringl-u-nnar	**?kringl-a** / ?kringl-na	kringl-a-nna / kringl-na-nna

Ayala and Hansson (2021) examined attestation rates of defectiveness conditioned by stem type and found that indeterminacy was a likely cause for defectiveness. Fw.OTHER lemmas were significantly less attested than Fw.N-FREE lemmas. The Icelandic Gigaword Corpus's (IGC; Steingrímsson et al. (2018)) lemma rates of attestation were then measured across three conditions: gender (Fw vs. Mw nouns), case (NOM.PL and ACC.PL vs. GEN.PL), and stem type. The first part of the study involved a categorical metric: for all the lemmas that were attested in the Fw or Mw classes in the NOM.PL and ACC.PL cases, how many did not have any attested GEN.PL tokens? The second part of the study used a continuous metric: was the rate of GEN.PL tokens significantly lower for Fw.OTHER type Fw nouns?

The categorical results suggested a pattern of attestation which confirms that indeterminacy is correlated with defectiveness. There were significantly lower rates of attestation for Fw.OTHER type lemmas, regardless of definiteness. Table 6 below shows the categorical results χ-squared test quantifying differences in underattestation under a metric termed "gappedness" (Ayala and Hansson 2021). What is most striking is that most lemmas are completely unattested in the GEN.PL; indefinite Mw lemmas were the best attested overall[3].

Table 6: χ-squared results of Fw vs. Mw lemma non-attestation (Ayala and Hansson 2021)

	Indefinite	Definite
-a stems	No significant difference	No significant difference
-(n)a stems	Fw (87%) > Mw (71%)	Fw (96%) > Mw (82%)

Ayala and Hansson (2021) also created a Poisson regression model of the same data, with individual tokens as an outcome variable, and overall lemma frequency, class, case, and stem type as predictors. The results for an indefinite set of tokens showed that despite there being fewer N-FREE type tokens overall, the interaction between stem type, case, and class indicated that OTHER type lemmas were significantly underattested. This effect was not seen, however, for definite tokens, which do not have a chance of being homophonous. This may have suggested that while indeterminacy leads to more categorical distributions of defectiveness, *both* indeterminacy and homophony appear to be factors for more gradient patterns of defectiveness over the Icelandic corpus.

2.2.3 Phonotactic ill-formedness

Sometimes, illegal consonant sequences result from inflecting consonant-final stems with the consonant-initial *-na* suffix. First, I will present three lemmas that end in a consonant cluster and do not exhibit defectiveness, and then we will look at two defective lemmas that end in a consonant cluster.

In table 7 below, three Fw words are presented whose stems end in a consonant cluster. The first lemma's stem (*hryðj-*) ends in a non-dorsal consonant + /j/, which indicates that is an N-FREE type. This word takes *-a* in the GEN.PL. The other two lemmas are both Fw.OTHER type, which means the suffix that they can take is not predetermined by their stem shapes. *Regla* ends in a stop + liquid sequence and can optionally take either *-a* or *-na*, while *kirkja* ends in a palatal stop and can only take *-na*. Since *hryðja* can only take *-a*, the sequence of consonants does not necessarily result in a complex onset or coda [r̥ɪ**ð.j**a], and so this sequence is fairly unproblematic. In REGLA, there is a similarly unproblematic syllabification with the *-a* variant [rɛk.la]. However, the *-na* variant will result in either a complex coda [rɛ**kl.n**a] or complex onset [rɛ**k.ln**a]. Crucially, in these first two lemmas, there are no stem alternations (NSA) that are triggered by *-na*, either because that suffix is not available (*hryðja*), or because the resultative complex cluster is accepted (*regla*).

Finally, *kirkja* is a Fw.OTHER type lemma that only takes *-na* in the GEN.PL and has a stem-final palatal stop. Since *-a* is not available, the NSA form is ruled out. However, palatal stops can only

[3] The exact counts will later be given in 4.

occur prevocalically, and so *[cʰɪ**rc.n**a] is ill-formed because *-na* is concatenated. Only the stem alternating (SA) variant where [c] alternates with [k] is well-formed: [cʰɪr**k.n**a]. These consonant stem alternations only ever occur with stem-final palatal stops.

Table 7: Stem shape alternations and stem-final consonant clusters

gloss	NOM.SG	GEN.PL *-a*	GEN.PL *-na* NSA	GEN.PL *-na* SA
hryðja 'downpour' (Fw.N-Free)	[r̥ɪðj-a]	**[r̥ɪðj-a]**	*[r̥ɪðj-na]	*[r̥ɪð-na]
regla 'rule' (Fw.Other)	[rɛkl-a]	**[rɛkl-a]**	**[rɛkl-na]**	*[rɛk-na]
kirkja 'church' (Fw.Other)	[cʰɪrc-a]	*[cʰɪrc-a]	*[cʰɪrc-na]	**[cʰɪrk-na]**

The defective paradigm in table 8 provides the forms for a Fw.Other lemma. Whereas the *-a* form is unavailable for *kirkja* in the gen.pl, speakers are much less certain about the availability of the *-a* form for fiðla. This paradigm is defective because in addition to the indeterminate *-a* form, the *-na* form is also unacceptable. This separates *fiðla* from *regla* and kirkja, since the latter two non-defective lemmas either have both suffixes available or only *-na*. However, the *-na* option results in either a complex coda ?[fɪ**ðl.n**a] or complex onset ?[fɪ**ð.ln**a]. The stem alternating variant *[fɪð.na] is ungrammatical, since stem alternations are only available for stems that end in a palatal stop.

Table 8: Defective paradigm for fiðla 'violin'

FIÐLA	SG.INDF	SG.DEF	PL.INDF	PL.DEF
NOM	**fiðl-a**	fiðl-a-n	fiðl-ur	fiðl-ur-nar
ACC	fiðl-u	fiðl-u-na	fiðl-ur	fiðl-ur-nar
DAT	fiðl-u	fiðl-u-nni	fiðl-um	fiðl-u-num
GEN	fiðl-u	fiðl-u-nnar	**?fiðl-na, ?fiðl-a, *fið-na**	?fiðl-na-nna, ?fiðl-a-nna, *fið-na-nna

Table 9 provides an example of a defective lemma that has a stem-final palatal stop *bylgj-* [pɪlc-]. Like *kirkja*, the *-a* form ?[pɪlca] is unavailable (although speakers are more uncertain about this availability for this lemma) and the *-na* form *[pɪlcna] is ill-formed. Unlike *kirkja*, however, the stem alternating variant ?[pɪlkna] is not completely acceptable either, even though this should be an available repair for stems that end in a palatal stop.

Table 9: Defective paradigm for bylgja 'wave'

BYLGJA	SG.INDF	SG.DEF	PL.INDF	PL.DEF
NOM	**bylgj-a**	bylgj-a-n	bylgj-ur	bylgj-ur-nar
ACC	bylgj-u	bylgj-u-na	bylgj-ur	bylgj-ur-nar
DAT	bylgj-u	bylgj-u-nni	bylgj-um	bylgj-u-num
GEN	bylgj-u	bylgj-u-nnar	**?bylg-na, ?bylgj-a**	?bylg-na-nna, ?bylgj-a-nna

2.2.4 Aversion to innovative allomorphy

Pertsova (2016) claimed that *lexical conservatism* (Steriade 1998) can contribute to an overall unwillingness to introduce novel allomorphs of a stem. Icelandic seems to exhibit a high degree of lexical conservatism, since most paradigm cells contain the same stems as consistent phonotactic units. This would entail that a learner would find it more economical to use existing forms rather than accept a new one. This principle may contribute to defectiveness, where suffixal allomorphy would not only motivate a learner's uncertainty toward a particular form, but would additionally violate an identity relation with the underlying stem. The other available form, *bylgja*, would conversely result in homophony with the NOM.SG, and so crucially, there is no unproblematic way to fill the GEN.PL cell of this paradigm.

The concatenation of the *-na* suffix invites another motivation for defectiveness because its CV shape can condition phonological alternations in consonant-final stems. These alternations can include vowel shortening {[kuːla]$_{\text{NOM.SG}}$, [kulna]$_{\text{GEN.PL}}$}, geminate shortening {[skemːa]$_{\text{NOM.SG}}$, [skemna]$_{\text{GEN.PL}}$}, consonant hardening {[stɔːva]$_{\text{NOM.SG}}$, [stɔːpna]$_{\text{GEN.PL}}$}, excrescent stopping {[viːsa]$_{\text{NOM.SG}}$, [vistna]$_{\text{GEN.PL}}$}, depalatalisation {[tʰɛːcʏr]$_{\text{NOM.PL}}$, [tʰɛhkna]$_{\text{GEN.PL}}$}, and preaspiration {[kaːta]$_{\text{NOM.SG}}$, [kahtna]$_{\text{GEN.PL}}$}. Lexical conservatism (Pertsova 2016; Steriade 1998) may prevent defectiveness if a potential form would require a base that has an acceptable lexical precedent. Sometimes, the phonological form of a word's morphosyntactic base would result in a form that would be phonotactically ill-formed; when this happens, a split-base effect allows a speaker to draw from multiple distinct references within the lemma's inflectional/derivational paradigm or even across multiple lexical paradigms. If no other reference is available to use as a phonological base, then defectiveness may occur when a speaker would be required to innovate a novel form with no corresponding lexical listing. For example, the word *kaka* 'cake' [kaːka] has a morphosyntactic base with the phonological form [kaːk]. If the *-na* suffix were appended to this base, then preaspiration would apply. However, in the KAKA lemma's inflectional paradigm, there are two phonological bases available*kak-* [kʰaːk] and *kök-* [kʰœːk], but there is no *[kʰahk] allomorph listed.

3 Research questions

1. Can the intuitions of native Icelandic speakers about where defectiveness occurs in noun inflection be confirmed in a large text corpus?

2. Does statistical underrepresentation correlate with lemma types that have been suspected to contribute to defectiveness? Namely, those that exhibit syncretism between the NOM.SG and GEN.PL, like Fw.N-FREE lemmas, and those that belong to the class of lemmas that are not uniformly defined by the affix they can take, like Fw.OTHER lemmas?

3.1 Predicted results

One of the main challenges of this work is that statistical underrepresentation is relative; a set of lemmas can only be underattested or unattested insofar as there must be some other set of lemmas against which a comparison can be made. There are two such datasets that will be compared in this study: a class-based set and a lemma-based set[4]. For each comparison between sets, there are different ways to interpret the factors, both individual and interacting, that might condition differences in attestation.

[4] As will be discussed in section 4.1, the mini-corpus has been curated such that NOM or ACC forms of lemmas are always attested; for this reason, along with the omission of the DAT case, case-based comparisons have not been considered.

In the class-based set, Fw lemmas are compared against Mw lemmas; this is a between-item design, since there is no overlap in Fw and Mw lemmas, and will help to identify any correlates of defectiveness that are unique to the Fw class. If the Fw class is defective, then we would expect that Fw lemmas would have significantly fewer GEN.PL tokens than Mw lemmas do. Additionally, this comparison also allows us to tease apart finer grained proxies for defectiveness that are distinguished by the properties of the Fw, but not the Mw. In particular, Fw lemmas have INDF.GEN.PL tokens that are syncretic with the INDF.NOM.SG when they are affixed by *-a*. For Mw lemmas, this identity relation does not exist for tokens affixed by *-a*. However, as seen in section 1, Mw INDF.GEN.PL forms that end in *-a* do exhibit syncretism across multiple cells in the paradigm. If avoidance of intraparadigmatic syncretism is a motivating factor for defectiveness, then it is expected that Fw and Mw lemmas should not be so differently attested unless identity between the NOM.SG and other parts of a paradigm is a particularly privileged kind of relation where distinctiveness should be maintained. Furthermore, Mw GEN.PL tokens can only be inflected with one allomorph, and so there should be no effect of stem type on the attestation of Mw lemmas, since only Fw could possibly be defective by way of uncertainty.

In the lemma-based set, a within-subject design is used to compare lemmas that belong to the same Fw class. In this set, case can be left out as a factor because such a comparison would be uninformative because the mini-corpus was specifically curated to ensure NOM.PL or ACC.PL attestation. By only looking within the set of Fw lemmas, we can infer how lemma-specific properties like stem shape can contribute to defectiveness in some paradigms but not others. For instance, if intraparadigmatic syncretism avoidance can be causally linked to defectiveness, then the following would be expected: since only indefinite tokens can be syncretic with the NOM.SG cell of the paradigm, it is expected that there should be fewer indefinite tokens for lemmas whose stem shapes select for the *-a* suffix than for stems that can obligatorily or optionally select for *-na*. In the same vein, if uncertainty causes defectiveness then it is expected that there should be fewer tokens for lemmas whose stem shapes can optionally select for *-na*, irrespective of definiteness.

3.2 Hypotheses

To summarise, there are three possible hypotheses (H_{1-3}) that can be inferred from the convergence of multiple factors; the visualisation of these factors is provided in the flowchart in Figure 1: intraparadigmatic syncretism avoidance, uncertainty, and the combination of the two. Figure 1 lists the expected causes of defectiveness resulting from potential underrepresentations in the GEN.PL of lemmas and tokens with certain properties.

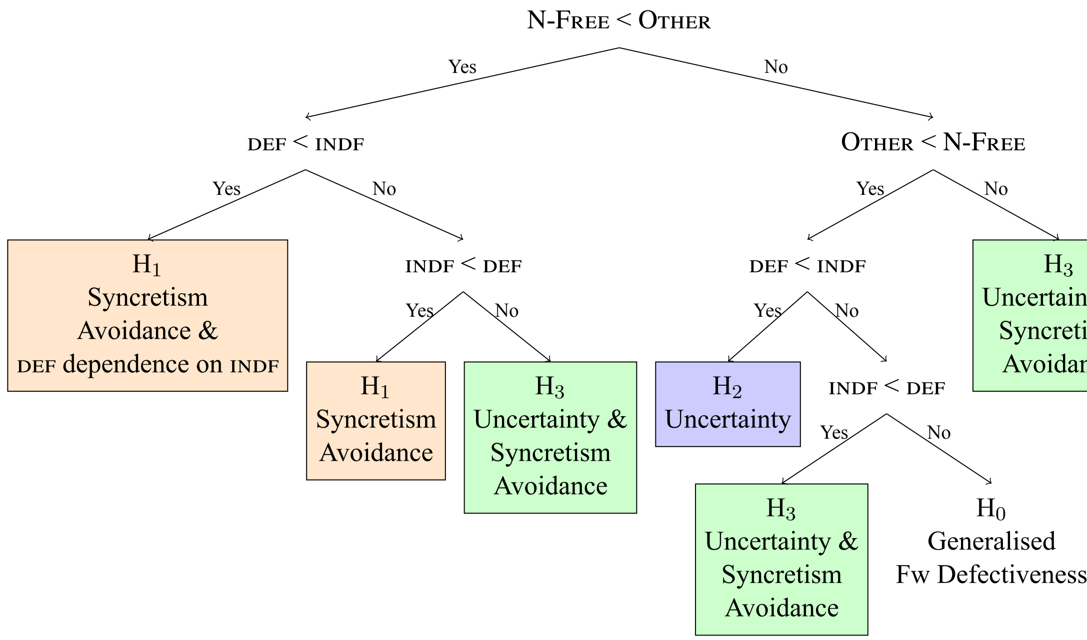


Figure 1: Inferences that can be made from different kinds of GEN.PL underrepresentation. Branching nodes ($x < y$) should be read as "x is more underattested than y".

3.3 Generalised Fw Defectiveness

Assuming that Fw lemmas are more underattested in the GEN.PL than Mw lemmas, the null hypothesis is that there are no great differences *within* the Fw inflection class. If there are no effects of class or definiteness on lemma attestation, then this study will have failed to have disproved this null hypothesis. This may suggest then that the GEN.PL cell of Fw lemmas is generally defective (e.g., by accident), or otherwise defective due to other factors that have not been explored here (e.g., ill formedness, semantic class).

3.4 Uncertainty

If Fw.OTHER lemmas are significantly more underattested in the GEN.PL than Fw.N-FREE lemmas, then this result can be interpreted as uncertainty-based defectiveness. The expected pattern of defectiveness would be as below in table 10, where a learner would be undecided between the *-a* and *-na* form of a lemma in both the INDF and DEF forms.

Table 10: Uncertainty in a lemma of stem type Fw.OTHER

HAMLA 'constraint'	INDF	DEF
NOM.SG	haml-a	haml-a-n
ACC.SG	höml-u	höml-u-na
DAT.SG	höml-u	höml-u-nni
GEN.SG	höml-u	höml-u-nnar
NOM.PL	höml-ur	höml-ur-nar
ACC.PL	höml-ur	höml-ur-nar
DAT.PL	höml-um	höml-ur-num
GEN.PL	**?haml-a / ?haml-na**	**?haml-a-nna / ?haml-na-nna**

3.5 Syncretism avoidance

There are two possible combinations of factors that could be interpreted as intraparadigmatic syncretism avoidance. The first type of syncretism avoidance, shown in table 11 below, will manifest where both class and definiteness are significant predictors of lemma attestation. Only INDF.GEN.PL tokens of Fw.N-FREE lemmas will be defective, since this cell is unambiguously homophonous with the NOM.SG.

Table 11: intraparadigmatic syncretism avoidance for an N-FREE lemma

LILJA 'lily'	INDF	DEF
NOM.SG	**lilj-a**	lilj-a-n
ACC.SG	lilj-u	lilj-u-na
DAT.SG	lilj-u	lilj-u-nni
GEN.SG	lilj-u	lilj-u-nnar
NOM.PL	lilj-ur	lilj-ur-nar
ACC.PL	lilj-ur	lilj-ur-nar
DAT.PL	lilj-um	lilj-u-num
GEN.PL	**?lilj-a**	lilj-a-nna

The second kind of syncretism avoidance can be inferred from a result where class may or may not be a significant predictor of lemma attestation, but where indefinite tokens of both Fw subclasses are underattested. In addition to the pattern shown in table 11, the pattern seen below in table 12 would also be predicted, since ostensibly *some* Fw.OTHER lemmas have the potential to end in -*a*. This form of syncretism avoidance interacts with uncertainty, since a learner is forced to select between two forms where one of the forms is homophonous with the NOM.SG.

Table 12: intraparadigmatic syncretism avoidance for an OTHER-type lemma

BYLGJA 'wave'	INDF	DEF
NOM.SG	**bylgj-a**	bylgj-a-n
ACC.SG	bylgj-u	bylgj-u-na
DAT.SG	bylgj-u	bylgj-u-nni
GEN.SG	bylgj-u	bylgj-u-nnar
NOM.PL	bylgj-ur	bylgj-ur-nar
ACC.PL	bylgj-ur	bylgj-ur-nar
DAT.PL	bylgj-um	bylgj-u-num
GEN.PL	**?bylgj-a** / bylg-na	bylgj-a-nna / bylg-na-nna

3.6 Uncertainty and syncretism avoidance

Both syncretism avoidance and uncertainty can be inferred to be at-play if there is both independent evidence for GEN.PL attestation of Fw.OTHER lemmas (tab. 10) and Fw.N-FREE lemmas (tab. 11). In addition to these independent motivations, certain Fw.OTHER lemmas may be susceptible to both uncertainty *and* syncretism avoidance. For example, KAKA (tab. 13) might show both uncertainty effects that would manifest in both the indefinite and definite GEN.PL, and it might show syncretism avoidance effects for indefinite *-a* forms in the GEN.PL. This differs from the uncertainty-only pattern in table 10, which would not be supported by independent evidence from underattestation in the GEN.PL of Fw.N-FREE lemmas. In that uncertainty-only case, indefinite *-a* forms in the GEN.PL could only be chalked up to uncertainty, and not syncretism with the NOM.SG.

Table 13: intraparadigmatic syncretism avoidance & DEF dependence for a lemma with the *-a* stem type

KAKA 'cake'	INDF	DEF
NOM.SG	**kak-a**	kak-a-n
ACC.SG	kök-u	kök-u-na
DAT.SG	kök-u	kök-u-nni
GEN.SG	kök-u	kök-u-nnar
NOM.PL	kök-ur	kök-ur-nar
ACC.PL	kök-ur	kök-ur-nar
DAT.PL	kök-um	kök-u-num
GEN.PL	**?kak-a / ?kak-na**	**?kak-a-nna / ?kak-na-nna**

3.7 Base dependence

If the Fw.N-FREE stem type is more underrepresented in the GEN.PL than the Fw.OTHER stem type, and definiteness is a significant predictor of GEN.PL underattestation, then this would mean that definite forms are dependent on the defectiveness of the indefinite base. The pattern shown below in table 14 is similar to the one in table 11, except that the definite cell is also expected to be defective. Notably, if class is not a significant predictor of attestation, then base dependence may be at work in both subclasses, resulting in a pattern that unites tables 13 and 14.

Table 14: intraparadigmatic syncretism avoidance in both the indefinite and definite forms of a Fw.N-Free lemma

LILJA 'lily'	INDF	DEF
NOM.SG	**lilj-a**	lilj-a-n
ACC.SG	lilj-u	lilj-u-na
DAT.SG	lilj-u	lilj-u-nni
GEN.SG	lilj-u	lilj-u-nnar
NOM.PL	lilj-ur	lilj-ur-nar
ACC.PL	lilj-ur	lilj-ur-nar
DAT.PL	lilj-um	lilj-u-num
GEN.PL	**?lilj-a**	**?lilj-a-nna**

4 Methods

Unlike with well-known cases of defectiveness in languages like Russian (cf. the dictionaries and grammars consulted by Pertsova (2016)) and Greek (cf. the online Greek dictionaries consulted by Sims (2015)), there is no Icelandic lexical resource (e.g., dictionary, word bank, etc.) that explicitly marks absent word forms[5]. All results in this study are inferred from positive evidence only in the Icelandic Gigaword Corpus (Steingrímsson et al. 2018), henceforth referred to as IGC. The methods described in this section will therefore pertain to how data were tidied and compiled as well as how the analytical models were created thereafter.

4.1 Data harvesting and tidying

The IGC (Steingrímsson et al. 2018) contains 1.55 billion automatically tagged words and is updated annually. The IGC's sources include web media (38%; e.g., blogs, forums), print (30%; e.g., books, newspapers), radio and TV (4%), official texts (26%; e.g., parliamentary speeches), and texts from the Árni Magnússon Institute for Icelandic Studies, the University of Iceland, and Icelandic Wikipedia (2%). The oldest texts date back to the 13th century, but 94% of the texts were written after 1980, and 86% after 2000. Tokens are lemmatized and tagged for gender, case, number, and definiteness[6].

For data harvesting, the IGC was first downloaded in its entirety and converted from XML format to text format. Elasticsearch[7] was used to index and create a searchable text database locally, which was then queriable for tokens using regular expressions[8]. The original corpus was organised into a hierarchical file structure (XML), with each file constituting a different text from a certain resource during a particular year. Each token was tagged between markup headers for lemma, word

[5] Defective GEN.PL forms, although not explicitly marked as missing, are sometimes listed in the Database of Icelandic Morphology (DIM; Bjarnadóttir (2021)) with both *-na* and *-a* alternatives. These forms are also sometimes annotated as "rare".

[6] These tags were relevant to the noun word category, but other subcategories for adjectives, articles, numerals, verbs, etc. (e.g., mood, voice, person, etc.) were also available. The entire MIM-GULL 1.0 Tagset is available from http://www.malfong.is/files/rmh_tagset_files_en.pdf

[7] https://www.elastic.co/

[8] Thank you to Miikka Silfverberg, who co-developed the IGC Corpus Reader: https://github.com/alexlilia/igc-corpus-reader

type, position in a sentence, and sentence number. These metadata were extracted for every token and concatenated into strings of three members in the following format `token0lemma0tag`, which were then indexed using Elasticsearch. Regex search queries could be used to look up any part of the `token0lemma0tag` string. The results of a search query were returned as an array with one column for hits and one column for contexts, where the hit was preceded and followed by seven tokens on either side.

The first step of the corpus curation process was automated. The initial search queries extracted tokens tagged as nouns that were either NOM.PL, ACC.PL, or GEN.PL. Gender was not directly searched for, but it was indirectly extracted using the queries for case, number, and definiteness below (since the exponents for these features differ by inflection class). Only disyllabic lemmas were extracted so that compounded forms were not accidentally included. Since words were not tagged for inflection class in the original corpus, certain restrictions on token and lemma shapes were given for each case-number combination to ensure that the nouns in the curated corpus belonged to the Mw or Fw class. They were as follows:

Mw Lemma (=NOM.SG form) ends in -i

- Mw INDF.NOM.PL Token ends in -ar
- Mw INDF.ACC.PL Token ends in -a
- Mw DEF.NOM.PLor DEF.ACC.PL Token ends in -arnir or -ana

Fw Lemma (=NOM.SG form) ends in -a

- Fw INDF.NOM.PL or INDF.ACC.PL Token ends in -ur
- Fw INDF and DEF.GEN.PL Token ends in -a
- Fw DEF.NOM.PL or DEF.ACC.PL Token ends in -ar

The second step of the corpus curation process was manual. As reported in Ayala and Hansson (2021), a random sample of data hand-checked by a native Icelandic speaker-linguist[9] revealed error rates up to 81.5%. Overall, errors were more common if (a) the lemma was very low-frequency, and (b) the wordform ended in *-a* (see appendix for error rates of Fw and Mw NOM-ACC.PL tokens).

These rates comprised false positives caused by tagging errors, lemmatisation errors, and noise (i.e., flotsam and jetsam). See Table 15 for a non-exhaustive list of false positive errors. Crucially, all the false positives included lemmas that ended in *-a*. For instance, within the tagging errors, verbs that ended in *-a* were often misidentified as Fw nouns, and other nouns from the strong masculine Mw inflection class and the weak neuter Nw class that all end in *-a* were wrongly identified as Fw nouns. Lemmatisation errors included real (though irrelevant) wordforms that were identified as belonging to a non-existent paradigm. For instance, both *fjölda* and *fola* are real wordforms, but they are irrelevant because they are Mw nouns whose lemmas should be FJÖLDI and FOLI, respectively. Wordforms that were misspelt, contextually misleading (e.g., a lack of capitalisation or punctuation led to proper nouns like *Helgja* being lemmatised incorrectly), or even non-Icelandic often resulted in non-existent lemmas. Spelling variations also led to a multitude of lemmas that ought to have been merged. Finally, there were noisy data that seemed to have been the result of poor OCR scanning

[9] Thank you to Gunnar Ólafur Hansson for hand-checking this mini-corpus.

(in the case of non-electronic texts) and non-linguistic word salads (e.g., hyperlinks, image hashes, keyboard mashing, etc.) in the case of electronic texts.

Table 15: Examples of errors culled during manual cleanup.

False positive error type	Error subtype	Lemma examples
Tagging error	Wrong word category	DÖKKNA$_{verb}$ 'to darken'; TWÖRKA$_{verb}$ 'to twerk'
	Wrong noun class	HERRA$_{Mw}$ 'lord, master'; SÉRA$_{Mw}$ 'reverend'; SKEMA$_{Nw}$ 'plan'; PRISMA$_{Nw}$ 'prism'; TEMA$_{Nw}$ 'theme'
Lemmatisation error	Real token, non-existent Fw lemma	FJÖLDA, FOLA, HÉDA, HÉDDA, HJERNA, HLIÐA, HLUTA, MYNDA, VEKJA, SKETSA, STUNDA
	Misspelt token, non-existent Fw lemma	BLIÐA, DOLGA, FÉLGA, FJÐL-SKYLDA, FULGA, GLIRNA, GREÐSLA, HELGJA, HERNA, JÖRÐA, LYJFA, NEFNDA, PLÖNTA, PÖNNA, RIKJA, SKEPPA, STOFNNA, STRIÐA, SÖGA, VONPA, VÖRNNA
	Spelling variations, multiple Fw lemmas	PITSA/PIZZA; PULSA/PYLSA
	Non-Icelandic token, non-existent Fw lemma	JEDA, LÖSA, PUDA, TRUDA, VESA, SKÖRA, ZDOLA, ZIMNA
Flotsam and jetsam	Non-existent token, non-existent Fw lemma	BAA, FMA, DSETA, FMA, GGJA, GJA, GRDA, GREA, HVFVETA, MTDA, ONA

Due to this, the present study could not be conducted on the whole corpus. Instead, a mini-corpus of indefinite and definite NOM.PL, ACC.PL, and GEN.PL tokens was hand-curated from Fw and Mw lemmas that were attested less than fifty times in either the NOM.PL or the ACC.PL of the indefinite set. The resulting mini-corpus contains 8262 tokens for 1056 lemmas. Since the lemmas in the GEN.PL only appear given that they were attested in either or both the NOM.PL or ACC.PL, the latter cases naturally have higher lemma and token counts. The mean (as well as median, min, and max) counts are given below in table 16.

Table 16: Summarised lemma attestation by class and case (in token counts)

class	case	mean	median	min	max
Mw	NOM.PL-ACC.PL	5.74	3	1	38
Mw	GEN.PL	0.75	0	0	22
Fw.N-Free	NOM.PL-ACC.PL	6.56	4	1	35
Fw.N-Free	GEN.PL	0.32	0	0	7
Fw.Other	NOM.PL-ACC.PL	6.09	3	0	36
Fw.Other	GEN.PL	0.19	0	0	21

There were many tokens missing in the GEN.PL, where the median was 0 for every class. As well, while Fw lemmas were on average, better attested than Mw lemmas in the NOM.PL-ACC.PL, they seem to be more underrepresented in the GEN.PL. To compare these rates, there were 169 Mw lemmas and 511 Fw that were not attested in the GEN.PL. The lemma and token counts are summarised below in table 17.

Table 17: Total number of lemmas and tokens in the mini-corpus by class and case

	NOM.PL or ACC.PL	GEN.PL
Mw	270_L; 2313_T	101_L; 303_T
Fw	598_L; 5456_T	87_L; 190_T

4.2 Frequentist models of attestation

This study employs two kinds of frequentist models that will be used to formulate predictions over the distributions of the class- and lemma-based sets from section 3: a categorical logistic regression and a gradient Poisson regression. The categorical model prods a more traditional conception of defectiveness where a paradigm contains a gap, such that a particular combination of morphosyntactic features are missing a realised form. The gradient model, on the other hand, captures the degree to which a lemma is attested. This is important for cases where external factors like register or medium may have motivated the usage of a form, but the form is still unacceptable to most speakers in most contexts. Since these two approaches are predicated on two different characterisations of defectiveness, they each address two sets of different, though sometimes overlapping, questions. Namely, a categorical model addresses the question of which forms are outright avoided, and the gradient model addresses questions about general trends of avoidance or hesitance for certain lexical items.

Two logistic regression models were created. For each model, the outcome variable was the probability that a lemma was attested (1) or unattested (0). There were three categorical predictors variables in the first model, which compared the Fw and Mw lemmas in the class-based set; these were stem type and definiteness. The next model compared stem type and case between three out of four combinations of class and definiteness; INDF.NOM-ACC.PL tokens would always be attested for both Fw and Mw lemmas, so these effects were not of interest.

Four separate Poisson regression models were created. For each model, the outcome variable was the number of tokens for a particular case/definiteness combination for a given lemma. There were three categorical predictor variables per model. These were stem type, case, and definiteness. There was also one continuous predictor variable, which was called *total*. This is the total number of tokens of that lemma (across all case/definiteness combinations) in the dataset. Aside from the always-attested lemmas in the combined INDF.NOM.PL and INDF.ACC.PL set, many lemmas were not attested for a given case/definiteness combination (i.e., token count=0).

In these models, the expected outcomes were counts that did not increase systematically with the total counts per lemma. This is expected because ideally, non-defective lemmas, whether frequent or infrequent, should exhibit similar proportions of token counts for a given case/number/definiteness combination. The results of the Poisson models should show that there is a consistent and regular increase of token count with the total count *if* there is nothing impeding token proportions from scaling.

One potential confound is that proportional attestation will always be higher for underpowered data sets because if the lemma is already fairly low frequency, then the attested forms will always have a higher proportion of tokens. Since Poisson regressions return proportional outcome variables (i.e., rates), an offset term for lemma frequency was included, intended to eliminate differences in frequencies by converting all counts to proportions. The problem with this is that it does not account for cases where a lemma is overall poorly attested and how that total frequency needs to be considered. Instead of an offset term, the workable fix in this study was to include the *lemma totals as a predictor*, and to be able to interpret interactions with token counts as changes in attestation over lemma frequency.

5 Results

import The results from one categorical (logistic regression) model and two gradient (Poisson regression) models will be presented in this section. These results will be dissected in three stages. First, I will present plots of the observed token counts per lemma with the expected values of the categorical (fig. 2) or gradient (fig. 3, 4) models overlaid on top of the data as regression lines. Next, this sketch of the data will be referred to when the significant effects are presented (tables 19-21) for each of the three models (summarised in table 18). Here, these three models will be compared for the similarities and differences between them. Finally, the predictors that are found to have significant effects will then be further unpacked for correlation, effect size, significance, and whether they interact with other factors.

The datasets used for these models determined how many levels the class and case factors included, with there being four models in all. There were three levels for class: Mw, Fw.Other, Fw.N-Free. These levels were Helmert contrast coded[10] for the following contrasts:

- [BC] the two Fw subclasses pooled against the baseline Mw class.
- [WC] the Fw.N-Free class was compared against the Fw.Other class.

The factor for case consisted of either a two-level predictor (nom.pl/acc.pl and gen.pl) with nom-acc.pl as the reference, or else all the data were characterised as gen.pl (gen.pl Only). Due to the complexity and number of continuous and categorical predictors, a benefit of simplifying these data was that it allowed for a maximum of three predictors (and their interactions!) per model. For each of the categorical and gradient approaches, lemma attestation/degree of attestation was the outcome variable predicted as a function of total **lemma frequency**, lemma **class**, and/or token **case**, and/or **definiteness**[11]. Both case and definiteness were also Helmert coded; these contrasts were centered around the means for nom.pl/acc.pl and indf, respectively.

[10] Weighted Helmert contrast regression coding by class (with each contrast centered around the means for Mw [BC] and Fw.Other [WC] based on the total number of observations per level to account for the differences in the amounts of lemmas between classes):

	Between Class (BC)	Within Class (WC)
Mw	-0.6862364	0.2148134
Fw.N-Free	0.3137636	0.7148134
Fw.Other	0.3137636	-0.2851866

[11] Definiteness is only included as a predictor in these models where case is excluded. Since def.gen.pl forms are not homophonous with the nom.sg, any homophony-motivated underattestation seen in the indf.gen.Pl should disappear when definiteness is involved as a factor. It is not very useful to model definite non-genitives.

Lemma frequency is included in every model to check if GEN.PL forms are underrepresented because their lemmas are also underrepresented overall, or if this underrepresentation indicates underattestation of GEN.PL tokens even across more frequent lemmas. Underattestation in Fw lemmas exhibits such a trend: as lemma frequency goes up, GEN.PL token frequency overall increases more slowly for the Fw classes than it does for the Mw class. These non-proportional increases in GEN.PL tokens more strongly support an interpretation of defectiveness, especially when such increases interact negatively with other predictors like definiteness.

The significant effects reported are summarised in section 5.2 in Table 18. Insignificant interactions were dropped and models were compared using likelihood ratio tests to achieve the most parsimonious models possible. It is important to note that all lemmas are only maximally attested by 60 tokens, so it should be maintained that these results might only be representative of low-frequency lemmas.

Overall, both the categorical and gradient models make similar predictions when case is removed as a predictor, and the data pool is reduced only to GEN.PL representation. The prediction these two model types have in common is that the lemmas in the Fw.OTHER class are underattested compared to the Mw. With the same GEN.PL data pool, the categorical and gradient models are both sensitive to definiteness as a significant effect when comparing the Fw.N-FREE subclass to the Fw.OTHER subclass. These findings, which will be elaborated below, overall suggest that the Fw are predicted to be underattested in the GEN.PL and less likely to be attested than the Mw lemmas, and that within the Fw class, Fw.OTHER lemmas are worse attested than Fw.N-FREE lemmas.

5.1 A sketch of the data and their models

Case is not an informative predictor for the categorical models since the data were curated to ensure all lemmas were attested in the NOM.PL/ACC.PL. The plot below (fig. 2) compares lemmas that are attested (n_{Mw}=125; $n_{\text{Fw.Other}}$=69; $n_{\text{Fw.N-Free}}$=27) against those that are unattested (n_{Mw}=279; $n_{\text{Fw.Other}}$=648; $n_{\text{Fw.N-Free}}$=138) in the INDF and DEF GEN.PL; the attested and unattested lemmas observed in the corpus are represented as dots and distributed along the y-axis at 0% (unattested) and 100% (attested), and the slopes overlaid on top of the graph are the logistic regression lines that represent the predicted probabilities of attestation at each given lemma frequency. With case removed as a predictor variable, class (BC; WC) and definiteness were all significant predictors of low probabilities of attestation, but no interaction effects were found.

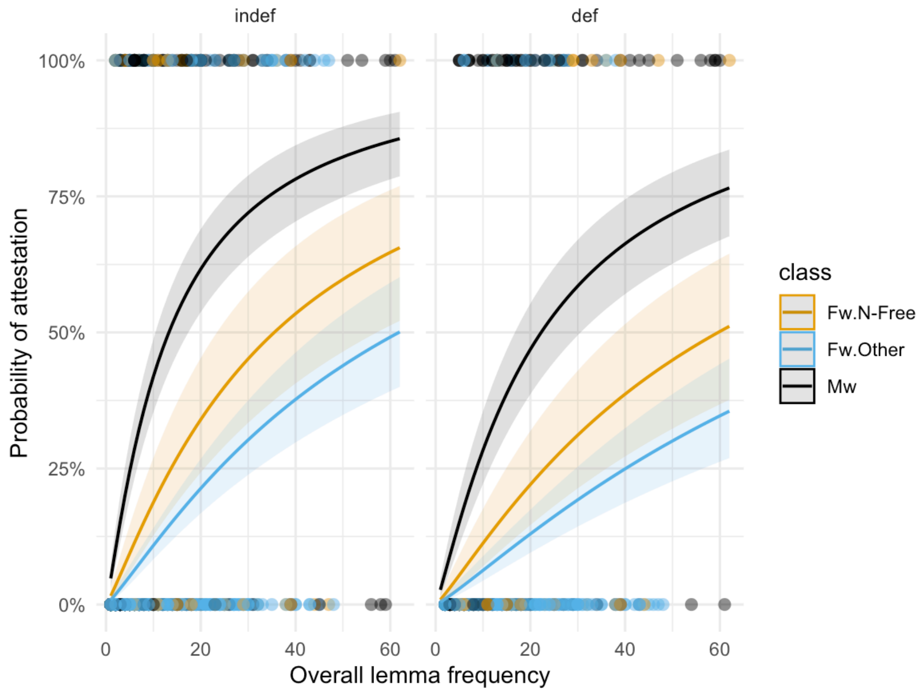


Figure 2: Categorical model of the predicted probabilities of INDF and DEF GEN.PL tokens

Since gradient models allow for gradated attestation, case is a more useful predictor variable here. The plot below (fig. 3) overlays Poisson regression lines overtop the observed lemmas represented by the plotted dots. In this model, where case is included as a predictor, we can observe that there are many more lemmas attested in the NOM-ACC.PL than in the GEN.PL, regardless of class. Case was found to be a significant predictor of lemma underattestation, both as a contrast and in interactions with lemma frequency and class.

Within the GEN.PL case, the two Fw subclasses are less attested than the Mw class. Only the WC contrast was found to have a significant negative interaction with case, meaning that Fw.N-FREE lemmas were better attested in the GEN.PL than Fw.OTHER lemmas. However, it is not clear that the degree to which Fw lemmas are underattested in the GEN.PL is greater than the degree to which Mw lemmas are underattested in the GEN.PL.

Of note, there were more Fw.OTHER lemmas than Mw and Fw.N-FREE lemmas. Despite this majority, the Fw.OTHER subclass was the worst attested in the GEN.PL case. This may be because there are more low-frequency (attested < 50 times) Fw.OTHER lemmas than there are of the other classes, which may speak to the Fw.OTHER class as an underrepresented class in general, but this is speculative without more data. One promising piece of evidence is that there is a significant three-way interaction between lemma frequency, the WC contrast, and case. Compared to the Fw.OTHER lemmas, Fw.N-FREE lemmas are worse attested with every increase of 1 in lemma frequency (both overall and in the GEN.PL).

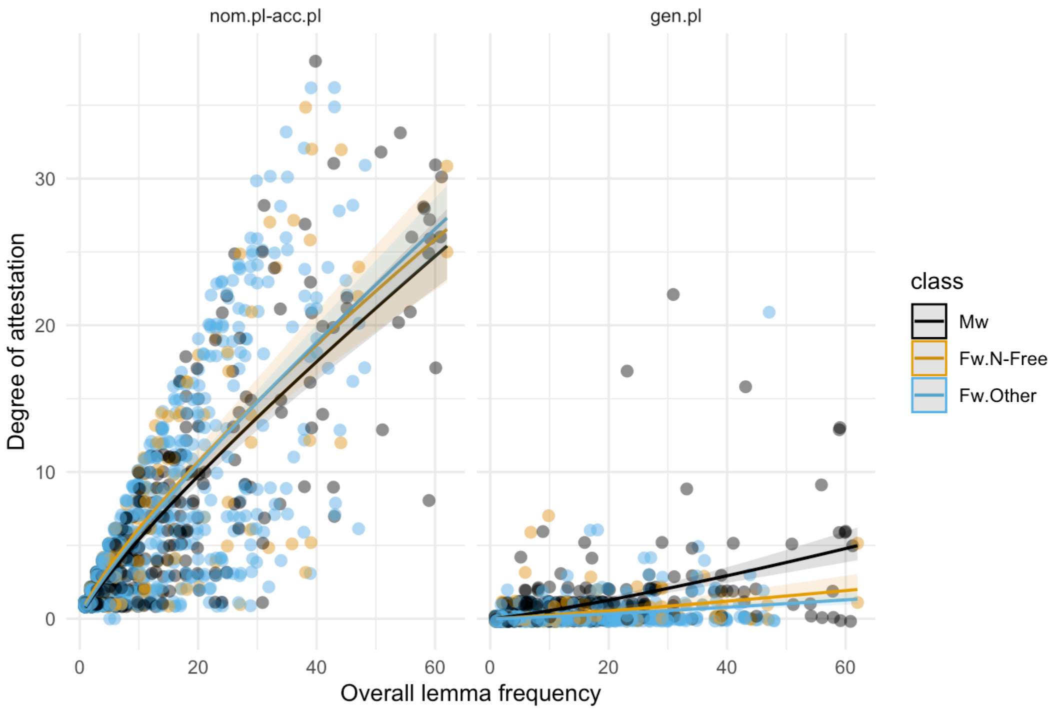


Figure 3: Gradient model of the predicted degrees of attestation of NOM-ACC.PL and GEN.PL tokens

Zooming in to lemmas that were only attested in the GEN.PL, the plot in figure 4 compares attestation of lemmas between the classes in the INDF and DEF. Definiteness, as a main effect, is found to significantly predict underattestation in all classes. Between the Mw and Fw classes, Mw lemmas are the best attested in the GEN.PL, regardless of definiteness.

Class (both the BC and WC contrasts) is found to significantly predict underattestation of the GEN.PL in the Fw subclasses. Between the two Fw subclasses, Fw.OTHER lemmas are worse attested in the GEN.PL than Fw.N-FREE lemmas; this is also supported by a contrast within the Fw class. With respect to definiteness, Fw subclasses are found to be more underattested in the DEF.GEN.PL compared to the Mw, which is evidenced by the interaction of class (BC) and definiteness. No such interaction was found within the Fw class, however; there is no significant difference between Fw.OTHER and Fw.N-FREE lemmas attested in the DEF.GEN.PL.

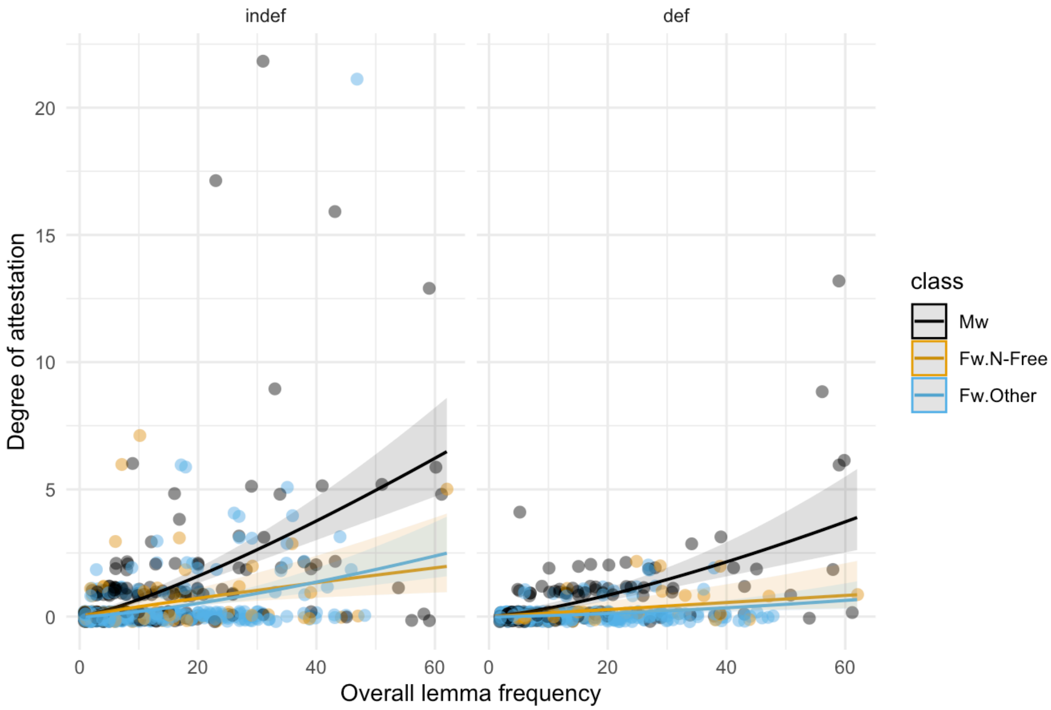


Figure 4: Gradient model of the predicted degrees of attestation of INDF and DEF GEN.PL tokens

5.2 Model comparison

I will summarise the ways in which the categorical and gradient models overlap, as well as in the ways that they differ. These models are condensed below in the effects summary table (18).

Table 18: A model comparison summary of all the significant predictors within models, whether these predictors correlated significantly independently or as an interaction term. The class-based data pool has three levels, where the data are Helmert coded where BC (Between Class) compares the means of the Fw subclasses against the mean of the Mw class and WC (Within Class) compares the mean of the Fw.N-FREE subclass against the mean of the Fw.OTHER subclass.

			*Significant Predictors (✓=main[12]; *=interaction)*				
	Model	**Available Cases**	**Freq**	**BC**	**WC**	**Case**	**Definiteness**
(18.a.i)[13]	Cat	NOM-ACC & GEN					N/A
(18.a.ii)	Cat	GEN Only	✓	✓	✓	N/A	✓
(18.b.i)	Grad	NOM-ACC & GEN	✓,*	*	✓,*	✓,*	N/A
(18.b.ii)	Grad	GEN Only	✓,*	✓,*	✓,*	N/A	✓,*

Overall, lemma frequency is a main effect in nearly every model (18.a.ii-18.b.ii) since it is expected that as lemma frequency increases, token frequency should also increase regardless of case/number or inflection class. The only exception for lemma frequency as a main effect is for the categorical model where case is included as a predictor (18.a.i), since the certainty of NOM-ACC.PL attestation ensures that nothing is significant in that model, and thus it has been left out here. This means that there is no way to directly compare the categorical and gradient models where case is included as a factor.

Where the data pool has been restricted to GEN.PL tokens only, the categorical and gradient models are by-and-large consistent: the expected outcomes of these two models confirm that, in the GEN.PL, Fw lemmas are on-average less likely to be attested and less attested than Mw lemmas, and that Fw.OTHER lemmas are less likely to be attested and less attested than Fw.N-FREE lemmas. These findings are expressed with class, both BC and WC, having significantly negative effects. The difference between these models that are characterised by only GEN.PL data is that the gradient model finds an interaction between class and definiteness, whereas the categorical model only has definiteness as a main effect. The expected outcome of the gradient model can be interpreted as the Fw subclasses being underattested relative to the Mw class in the DEF.GEN.PL compared to the INDF.GEN.PL. With a binary outcome variable that is only concerned with whether a lemma is attested or not, this nuance is lost in the categorical model.

Finally, when case is not included as a predictor, the gradient models show that the BC class contrast has a significant interaction with definiteness. Both gradient models, whether case is included or not, show that WC class contrasts have significant interactions with lemma frequency. When case is included as a predictor, WC class contrasts have significant interactions with case; i.e., the underrepresentation of GEN.PL as compared to NOM-ACC.PL is stronger (the attestation gap between them is bigger) in the Fw than in Mw, and in Fw.OTHER than in Fw.N-FREE.

Fw lemmas are worse attested than Mw lemmas and Fw.OTHER lemmas are worse attested than Fw.N-FREE lemmas in the GEN.PL compared to the NOM-ACC.PL.

5.3 A deeper dive into the models' effects

Moving away now from the model comparisons, this discussion will now centre on the predictions of the significant effects in each model. Specifically, the following effects: (1) frequency almost always has a significant main effect (to the exception of the categorical models with case as a factor), but is never found to interact with class; (2) compared to the Mw, Fw lemmas are less attested in the GEN.PL, and (3) DEF.GEN.PL tokens are better attested within Mw lemmas than Fw lemmas. Due to the choice of Helmert coding for class, it should be noted that the reference level for class contrasts depends on whether the predictor is labelled as class [BC] or class [WC]: the effects involving class [BC] should be interpreted as the differences between the (Fw – Mw) classes, and class [WC] as the difference within the (Fw.N-FREE – Fw.OTHER) subclasses.

First, we will look at the categorical GEN.PL Only model (see figure 2; model 18.a.ii). These categorical models are logistic regressions, and their coefficients are given as log-odds. The ex-

[12] For BC and WC, these are not main effects, but rather they are contrasts for the class predictor. When these contrasts are independently significant, they are marked with a checkmark, otherwise they are marked with an asterisk if the contrast(s) are significant in an interaction term.

[13] As expected, no effects were found for model (18.a.i), compared to the Mw NOM/ACC reference, since the NOM-ACC was attested for all lemmas 100% of the time.

ponentiated coefficients return the odds ratio, which are the (proportional) increase or decrease in the odds that a lemma is likely to be attested rather than not. With every increase of 1 in lemma frequency, there is a 76% chance that a lemma will be attested in the GEN.PL, regardless of class. Fw lemmas are much less likely than Mw lemmas to be attested in the GEN.PL. The significant contrast effect of class [BC] shows that there is only a 19% chance a lemma will be attested in the Fw class. Since there are no interaction effects found in this model, it should be noted that this effect, as well as class [WC] and def [DEF] exist independently of one another. Thus, Fw lemmas are shown to be underattested in the INDF.GEN.PL irrespective of lemma frequency. Furthermore, the effect of class [WC] shows that Fw.N-FREE lemmas have a 65% chance of being attested in the GEN.PL, regardless of definiteness. Overall, lemmas only had a 35% chance of being attested in the DEF.GEN.PL.

Table 19: Categorical model of the predicted probabilities of INDF and DEF GEN.PL tokens

	Attested		
Predictors	*Log-Odds*	*CI*	*p*
(Intercept)	-4.32 ***	-4.86 – -3.82	**<0.001**
lemmaFreq	1.16 ***	0.97 – 1.35	**<0.001**
class [BC]	-1.46 ***	-1.83 – -1.10	**<0.001**
class [WC]	0.64 *	0.11 – 1.15	**0.016**
def [DEF]	-0.60 ***	-0.96 – -0.25	**0.001**
Observations	1286		
R^2 Tjur	0.234		

* *p<0.05* ** *p<0.01* *** *p<0.001*

Next, in the gradient models, we will first look at the results where case is included as a factor (see figure 3; model 18.b.i). Since the NOM-ACC.PL case is included, there is much more data available to sample, and so the R2 value for this model is much higher than those seen in tables 19 and 21. Also, unlike in the tables of the categorical results (which are given as log-odds), the gradient results are given as log-means. These results are therefore not the likelihoods that a lemma will be attested, but rather to what degree lemmas are more, or less, attested given the presence of an effect.

In summation, the findings show that lemmas are predicted to be underattested in the GEN.PL compared to the NOM-ACC.PL, and that the degree of this GEN.PL underattestation differs within Fw subclasses (Fw.OTHER < Fw.N-FREE). Every increase of 1 in lemma frequency is predicted to result in a 74% chance of lemma attestation. Lemmas only have a 2% chance of attestation in the GEN.PL compared to the NOM-ACC.PL, regardless of class or lemma frequency. When interacting with lemma frequency, however, lemmas had a 60% increase in attestation (for every increase of 1 in lemma frequency) in the GEN.PL compared to the NOM-ACC.PL. The interaction of case and class [WC] shows that the Fw.N-FREE class has an 87% of attestation in the GEN.PL compared to the Fw.OTHER. However, when these factors are involved in a three-way interaction with frequency, the chance of FwN-FREE attestation drops to 38% with every increase of 1 in lemma frequency.

Table 20: Gradient model of the predicted degrees of attestation of NOM-ACC.PL and GEN.PL tokens

	Degree of Attestation		
Predictors	*Log-Mean*	*CI*	*p*
(Intercept)	-2.24 ***	-2.49 – -2.01	**<0.001**
lemmaFreq	1.04 ***	0.96 – 1.11	**<0.001**
class [BC]	-0.22	-0.60 – 0.15	0.251
class [WC]	1.04 ***	0.43 – 1.62	**0.001**
case [GEN.PL]	-4.10 ***	-4.59 – -3.64	**<0.001**
lemmaFreq * class [BC]	-0.09	-0.21 – 0.03	0.124
lemmaFreq * class [WC]	-0.27 **	-0.47 – -0.07	**0.007**
lemmaFreq * case [GEN.PL]	0.39 ***	0.24 – 0.54	**<0.001**
class [BC] * case [GEN.PL]	-0.70	-1.46 – 0.04	0.069
class [WC] * case [GEN.PL]	1.87 **	0.67 – 3.05	**0.002**
(lemmaFreq * class [BC]) * case [GEN.PL]	-0.16	-0.40 – 0.08	0.191
(lemmaFreq * class [WC]) * case [GEN.PL]	-0.49 *	-0.88 – -0.09	**0.014**
Observations	2572		
R^2 Nagelkerke	0.997		
	* *p<0.05*	** *p<0.01*	*** *p<0.001*

Finally, in this gradient model where case has been removed as a factor (see figure 4; model 18.b.ii), we see that most of the predictions concerning degrees of attestation are consistent with the predictions over attestation seen in table 19. With every increase of 1 in lemma frequency, lemmas have a 79% chance of being attested in the GEN.PL. All lemmas, regardless of class or frequency, are also predicted to only have a 19% chance of being attested in the DEF.GEN.PL. Fw lemmas were found to be only 30% likely to be attested attested in the GEN.PL(controlling for lemma frequency), and 36% likely to be attested in the DEF.GEN.PL, the latter of which was an interaction not found to be significant in the categorical model. This gap suggests that these DEF.GEN.PL tokens are not outright missing. This interaction results in a very weak effect, however; in figure 4, it seems as though the trajectories of the Fw.OTHER and Fw.N-FREE slopes seem to switch between the indefinite and definite subplots. This would be an interesting effect to further probe in future work with more data. Fw.N-FREE lemmas were found to have a probability of 88% in the INDF.GEN.PL, but there was no interaction found with definiteness.

Table 21: Gradient model of the predicted degrees of attestation of INDF and DEF GEN.PL tokens

	Degree of Attestation		
Predictors	*Log-Mean*	*CI*	*p*
(Intercept)	-4.56 ***	-5.08 – -4.09	**<0.001**
lemmaFreq	1.31 ***	1.16 – 1.47	**<0.001**
class [BC]	-0.83 *	-1.58 – -0.10	**0.029**
class [WC]	1.96 **	0.76 – 3.12	**0.001**
def [DEF]	-1.47 **	-2.43 – -0.55	**0.002**
lemmaFreq * class [BC]	-0.10	-0.34 – 0.13	0.387
lemmaFreq * class [WC]	-0.49 *	-0.87 – -0.11	**0.012**
lemmaFreq * def [DEF]	0.11	-0.16 – 0.40	0.431
class [BC] * def [DEF]	-0.56 *	-1.05 – -0.10	**0.018**
class [WC] * def [DEF]	0.49	-0.33 – 1.27	0.232
Observations	1286		
R^2 Nagelkerke	0.635		

* *p<0.05* ** *p<0.01* *** *p<0.001*

To summarise these results, both the categorical and gradient models predict that lemmas are underrepresented in the GEN.PL compared to the NOM-ACC.PL, as well as in the INDF.GEN.PL compared to the DEF.GEN.PL, and that the lemmas in the Fw class are unattested/underattested compared to the Mw class, and that the lemmas in the Fw.OTHER subclass are unattested/underattested compared to the Fw.N-FREE subclass. However, Fw.N-FREE lemmas appear to have slower increases in attestation proportional to lemma frequency compared to increases in Fw.OTHER lemmas. A final weak result in the gradient model seen in table 21 (figure 4) is that there seem to be fewer Fw lemmas attested in the DEF.GEN.PL than Mw lemmas, but there is no such interaction found within the Fw class. Since there are already so few definite tokens, and the GEN.PL Only models have lower R2 values, it seems possible that with more datapoints, this area of the data may yield interesting divergence in the behaviour of Fw lemmas in the DEF.GEN.PL.

6 Discussion

This part will address the research questions presented in section 3. To do so, we will revisit the results in section 5 and forge a clearer link with the conclusion that uncertainty (6.1), syncretism avoidance (6.2), and base dependence (6.3) are all driving defectiveness. For each of these drivers, I will propose morpho-phonological loci for failure where a learner's uncertainty blocks word formation. There are currently no theoretical models that can appropriately undergenerate forms for the Icelandic Fw lemmas without creating gaps elsewhere in the Fw inflection class, but I propose that existing models can be set up to crash This interpretation will then lead into a brief review of the limitations of this study, and how the experiment design could be adjusted in the future when more data become available.

At least for low-frequency lemmas (those attested less than fifty times), the categorical and gradient models in section 5 confirm that the defective GEN.PL is specifically underattested in the Fw inflection class. While it is possible that there are other motivations for defectiveness that affect only the GEN.PL cell of a paradigm, it seems unlikely to me that this defectiveness would only occur within the Fw and not the Mw inflection class. There are no notable semantic divisions that I am aware of between these two classes that could make the genitive and/or plurality more or less felicitous[14].

As shown in section 4, one of the biggest challenges of this work has been to define what "underattestation" means without access to direct negative evidence. Therefore, it bears repeating that the "underattestation of Fw lemmas in the GEN.PL" is a conclusion drawn from the following relative measures: it is true that for both the Fw and Mw inflection classes lemmas are better attested by NOM.PL or ACC.PL tokens than GEN.PL tokens. However, the degree to which lemmas are more poorly represented by GEN.PL tokens is significantly greater for Fw lemmas than Mw lemmas. This underrepresentation was predicted in model 18.b.i, which shows that case has a significantly greater effect on underattestation when it interacts with class.

As mentioned in section 5, this interaction could only be probed in the gradient models. Due to the nature of the data curation, where NOM.PL or ACC.PL tokens were categorically present, there were no effects for the categorical model (18.a.i) that could predict whether Fw lemmas were somehow "gappier" (more categorically unattested) than Mw lemmas in the GEN.PL case. To explore such a question, one would require a dataset where the reference is not NOM.PL or ACC.PL. One potential avenue for analysing this hypothetical dataset would be to examine attestation at the level of the lemma, rather than the inflection class. With a within-lemma design, gappedness could be measured in relation to expected non-attestation of *other lemmas* rather than other inflection classes. The benefit of this design would be that the attestation would be measured relatively between all case-number-definiteness combinations, rather than relative to a single case-number-definiteness combination, as in this study. For example, in figure 5 below, the degree of attestation of a lemma in the GEN.PL is given as a function of that same lemma's degree of attestation in the NOM-ACC.PL.

[14] The Nw class is much more semantically restricted, and is often limited to words concerning body parts that come in pairs (e.g., *auga* 'eye', *eista* 'testicle', *nýra* 'kidney', etc.). It would be interesting to compare the incidence of GEN.PL tokens between the Fw and Nw in this case, since it seems that plurality is more inherent to the Nw class than the Fw class, and it is not clear whether this inherency would advantage the Nw GEN.PL.

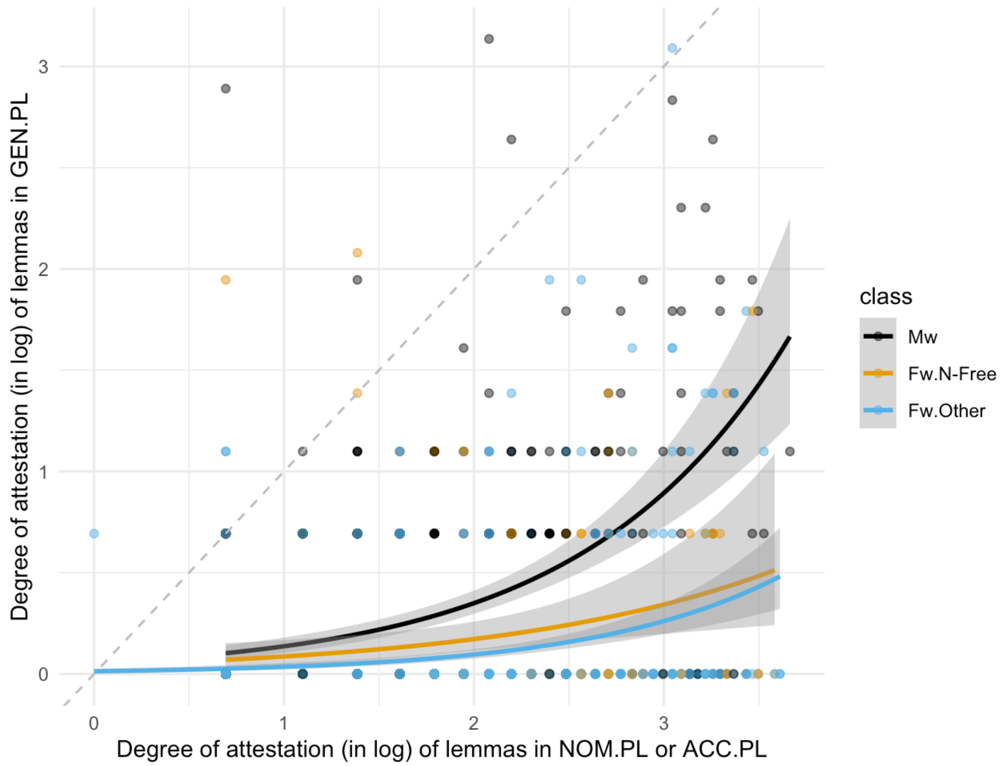


Figure 5: Degree of attestation (in log) between cases by lemma

The Poisson regression curves in this plot predict that the GEN.PL is better proportionally represented in Mw lemmas than in Fw lemmas. This method could then be applied for a wider range of cases, where no case-number-definiteness combination would need to be the always-attested reference. Each lemma could then be analysed for trends such as phonological [stem] shape, which is more concrete than the stem shape classes defined in this study. Additionally, figure 5 shows that as lemmas become more frequent, this trend becomes more obvious. More data would favour this approach, particularly the lemmas excluded from this study with frequencies greater than fifty. With these considerations in mind for future work, the current gradient finding (model 18.b.i) is sufficiently consistent with the reported intuitions that defectiveness is specific to the GEN.PL of Fw nouns, which thus confirms the first research question.

Now we will move on to the categorical and gradient models 18.a.ii and 18.b.ii, which restrict the datapool to GEN.PL tokens only and introduce definiteness in-place of case as a predictor. Both models predict that the contrast within the Fw class is significant, and that lemmas are less likely to be attested/are more underattested in the Fw.OTHER subclass than the Fw.N-FREE subclass. Definiteness was only involved in a significant interaction with the between-class contrast in the gradient model (see table 4), but not the categorical one (see table 2). Therefore, the models predict that definiteness is a stronger predictor of underattestation for Fw lemmas than Mw lemmas; there is no such finding *within* the Fw class. The second research question was whether there are any correlations between underrepresentation and morpho-phonological characteristics. Since Fw.OTHER lemmas are more poorly attested in the GEN.PL, and Fw.OTHER lemmas comprise an inflection class for which the choice of GEN.PL exponent is not categorically predictable, this model suggests that uncertainty drives defectiveness.

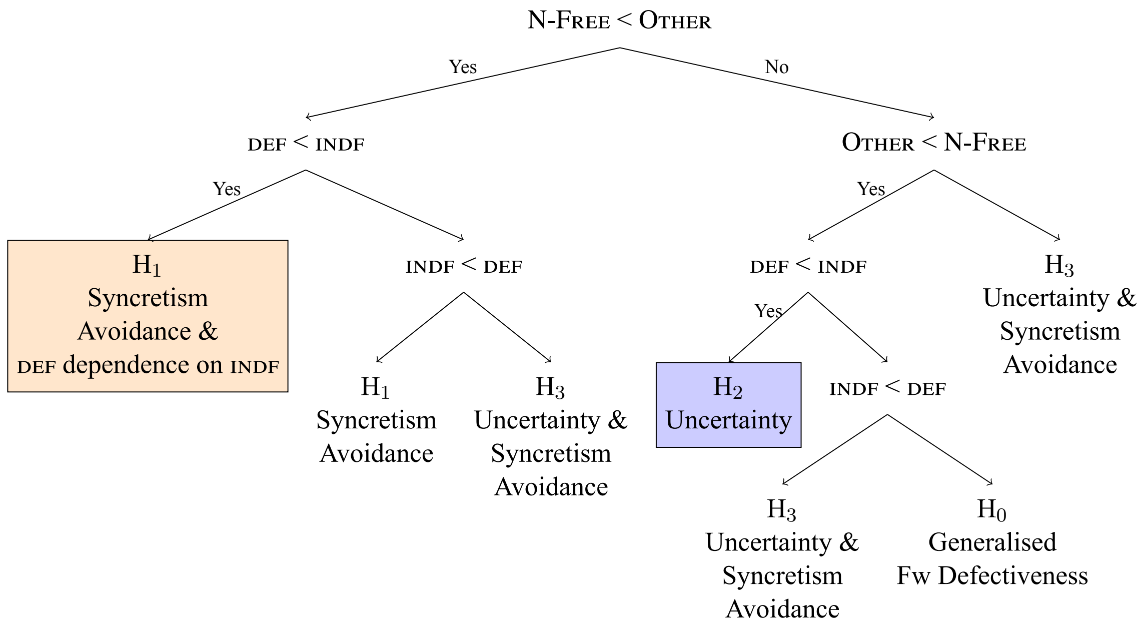


Figure 6: Summarised results of the models

If it were the case that syncretism avoidance alone caused the defectiveness seen in this dataset, then the interaction between the within-class contrast and definiteness should have been significant in the gradient model. This is because the encliticisation of the definite article, undoing the nom.sg = gen.pl syncretism, would have been expected to improve attestation of Fw.N-Free lemmas, since defectiveness would have been expected to occur had syncretism avoidance been the main source. Unfortunately, this interaction was not significant, and there were furthermore no significant interactions found in the categorical model either, and so the null hypothesis was not rejected. Interestingly, there was an effect found for the interaction of the between-class contrast and definiteness, where definiteness corresponded with Fw lemmas being only 36% likely to be attested in the def.gen.pl. Thus, across the entire class of Fw lemmas, there seemed to have been many fewer def.gen.pl tokens than in the Mw class.

There are three possible reasons for this difference: the first is that if a word form is defective in the indf, it is less likely that a learner will produce its def counterpart. This explanation applies regardless of whether uncertainty or syncretism avoidance caused defectiveness in the indf form and assumes that there is no difference between the Fw.Only and Fw.N-Free. The second possibility is that there is a discrepancy in attestation between the Fw.Only and Fw.N-Free, and that this discrepancy is not detectable because of a lack of data, but that it is sufficient to cause underrepresentation of the Fw compared to the Mw. The third possibility is that both the first and second possibilities are true, and that Fw.Other lemmas are underattested in the indf.gen.pl, thus provoking underattestation in their definite counterparts, in addition to syncretic Fw.N-Free indf.gen.pl tokens provoking underattestation in their definite counterparts. Since syncretism avoidance has not been altogether ruled out, I will incorporate a discussion of a potential theoretical mechanism that can account for both the intraparadigmatic awareness necessary in addition to incorporating a model that can explain why definite forms may be dependent on the existence of their indefinite bases.

These results establish non-accidental factors for the patterns of defectiveness that are both observed and reported in Icelandic nominals. These factors point to uncertainty as a primary cause of defectiveness, with syncretism avoidance as a potential secondary cause. These sources of defectiveness must have been learnt at some point in the grammars of Icelandic speakers, though not so severely as to result in absolute gaps. We know this fact because the gradient models in this study have enabled observation of finer grained differences in attestation that were lost in the low resolution categorical models. In order for these patterns to have been acquired, there need be necessary theoretical mechanisms that are associated with these causes, but as of the present time, there is no any one mechanism that exists. In this section, I will implement two theoretical models that may account for both defectiveness and syncretism avoidance, but it is important to note that these are not the only models available[15].

6.1 Uncertainty as a mechanism for defectiveness in Fw.Other lemmas

Albright (2003) first modelled uncertainty in Spanish paradigm gaps using the Minimal Generalisation Learner (henceforth referred to as MGL; Albright and Hayes (2003)). The principle in Albright's experiment was that morphophonological uncertainty could be measured as the likelihood that a morphophonological rule applies, which he terms the "reliability" of the rule. The MGL takes an input list of form pairs (namely INF ~ 1SG in Albright (2003)), and induces morphophonological rules from the surface differences between pairs. This rule consists of (a) the observed structural change and (b) the phonological environment in which the change occurs. The reliability of a rule is then based on how frequently a change occurs in a given environment. A reliable rule is one where a change is almost always expected to occur in a given phonological environment. A learner's uncertainty about a form should therefore increase as a rule's reliability decreases.

We can apply this model to Icelandic. For the purposes of this discussion, I will illustrate the MGL approach with the small training dataset in table 22. However, outside the scope of this qualifying paper, the results would ideally be calculated over the entirety of the Mw and Fw sets.

[15] Furthermore, the implementation of both of these models means that I am purporting that there are two distinct sources of defectiveness. However, these defectiveness**es** were not analysed separately in this study! It is totally possible that one source *does* undergenerate forms to the point of absolute gappedness, but that this distinction has been lost in the confluence of the two sources. It would also be interesting to know if these two sources interact, since concretely, the selection of a syncretic form could either facilitate or impede a learner's uncertainty between a *-a* or *-na* form. On the one hand, a learner may start out with a *-na* form perfectly, and therefore be able to rule out the syncretic *-a* form easily, thus eliminating uncertainty. On the other hand, the learner may be starting with an illicitly syncretic *-a* form as a reference point, and may be *introduced* to uncertainty when a novel *-na* form presents as a repair.

Table 22: Sample training input pairs for the MGL

class	INDF.NOM.SG	INDF.GEN.PL	gloss
Mw	[skrahtɪ]	[skrahta]	*skratti* ‘demon’
Mw	[lahpɪ]	[lahpa]	*lappi* ‘laptop’
Fw.N-Free	[sau:na]	[sau:na]	*sána* ‘sauna’
Fw.N-Free	[flʏntra]	[flʏntra]	*flundra* ‘flounder’
Fw.Other	[kru:pa]	[kru:pa]	*grúpa* ‘group’
Fw.Other	[svɪ:pa]	[svɪhpna]	*svipa* ‘riding crop’

Crucially, the role of the MGL is to consolidate the above into the most generalised rules possible. To do so, we must start at “the location of change” (Albright and Hayes 2003:124) and retain any segments that overlap between phonological environments. If there are no segments that overlap in the *immediate environment* of the location of change, then features are used instead. Any other mismatches are converted into a variable. For example, the rules in example 4 do not share segments immediately to the left of the location of change. However, both [ht] and [hp] are voiceless preaspirated stops, and so this may be generalised to a natural class. [a] overlaps in both words, and the rest {*skr*, *l*} can be reduced to a single variable X. The new generalised rule is provided below in example (4).

Starting with the Mw forms, we see the following structural change of ɪ → a after the stems {skraht-, va:v-} to form the GEN.PL.

(4)

Observed alternation	Generalised rules	
ɪ → a / [skraht__]$_{[\text{GEN, PL}]}$ ɪ → a / [lap__]$_{[\text{GEN, PL}]}$	ɪ → a / X[-son, -cont, +s.g.]_$_{[\text{GEN, PL}]}$	(Mw -a rule)

In the forms that belong to the Fw.N-Free subclass, there are effectively no structural changes (a → a) that take place to form the GEN.PL. There are no shared segments immediately adjacent to the location of change, but [n, r] do form a natural class of sonorants. The remaining material {sau:, flʏnt} is reduced to a single variable X. The new generalised rule is shown below in (5).

(5)

Observed alternation	Generalised rules	
a → a / [sau:n__]$_{[\text{GEN, PL}]}$ a → a / [flʏntr__]$_{[\text{GEN, PL}]}$	a → a / X[+cons, +son]_$_{[\text{GEN, PL}]}$	(Fw.N-Free -a rule)

Finally, in the Fw.Other subclass, we see that there are two strategies for GEN.PL formation: either effectively no change (a → a), or a (a → na) rule. Unlike for the tokens that belonged to Mw or Fw.N-Free lemmas, two generalised rules must be created for the specific rules in (6) because there are two different structural changes that can occur in the same environment.

(6)

Observed alternation	Generalised rules	
a → a / [kru:p__]$_{[\text{GEN, PL}]}$	a → a / Xp_$_{[\text{GEN, PL}]}$	(Fw.Other -a rule)
a → na / [svɪ:p__]$_{[\text{GEN, PL}]}$	a → na / Xp$_{[\text{GEN, PL}]}$	(Fw.Other -na rule)

Here we notice that in examples (5, 6), the same changes occur in their respective phonological environments 100% of the time[16]. However, in example (6), we see that the structural changes each only occur 50% of the time in the same phonological environment. In summation, the rules that are induced from tokens in the Mw and Fw.N-Free classes are more **reliable** than the rules that are induced from the tokens in the Fw.Other class. In practice, this means that if a learner were given a test input like [kli:pa]$_{[\text{NOM, SG}]}$ (*klípa* 'difficulty'), then there would be a 50% chance it would be inflected as [kli:pa]$_{[\text{GEN, PL}]}$ and a 50% chance it would be inflected as [klihpna]$_{[\text{GEN, PL}]}$[17].

The training data seen in examples (4 - 6) mostly show concatenative processes. However, oftentimes, concatenation of *-na* can also trigger a change in the stem. For instance, the input pair [svɪ:pa] ~ [svɪpna] also involves vowel shortening in the closed syllable of the gen.pl form. A benefit of using the MGL is that it is agnostic to morphological boundaries when it learns a morphophonological change. A possible addition to the *-na* rule in example (6) could be the more specific rule seen below in (7).

(7)

Observed alternation	Generalised rules	
a → a / [kru:p__]$_{[\text{GEN, PL}]}$	a → a / Xp_$_{[\text{GEN, PL}]}$	(Fw.Other -a rule)
a → na / [svɪ:p__]$_{[\text{GEN, PL}]}$	a → na / Xp_$_{[\text{GEN, PL}]}$ V:pa → Vhpna / X__ $_{[\text{GEN, PL}]}$	(Fw.Other -na rule i) (Fw.Other -na rule ii)

A rule like (7) would require that learners would end up generalising two sets of rules: those for stems that end in a consonant, and those that end in a vowel. The existence of two sets of rules removes the problem of uncertainty, but it introduces two issues: another kind of uncertainty, and a very narrow pattern.

First, the rules for stems that end in a consonant rely heavily on specific structural changes, while the changes for the stems that end in a vowel are much broader, though their phonological contexts are more specific. This means that a learner must decide which set of rules is more reliable, because otherwise, words like [svɪ:pa] would simply enforce a zero-derivation strategy according to the first generalised rule in (6). Thus, it is possible that "uncertainty" is not necessarily a source of defectiveness that forces a learner to choose between multiple structural changes in the same phonological environment, but it could alternatively refer to a learner's uncertainty over which environment to choose.

Second, Albright (2003) discusses how frequency and uncertainty interact, where he theorises

[16] This is an extremely optimistic subset of the data, and we would of course not expect to find naturalistic data this tidy.

[17] Keeping in line with the rest of this qualifying paper, I am only considering variation as a property of the inflection class, and not of the lemmas themselves. That being said, BÍN reports both the *-a* and *-na* variants for all three Fw.Other words provided in these sample training and testing data!

that if a word is unfamiliar or infrequent and there is “no high-reliability ‘default’ pattern” (p. 10) to fall back on, then there are two problems which arise: the learner has no access to any memorised word forms, nor does she have any reliable rules that she can generalise from other parts of the inflection class. The latter of which is either caused by low-reliability patterns or highly specific patterns that cannot be extended to a novel lexical item. In the Icelandic data, if a learner is uncertain about whether to apply (a → a) or (a → na) (7), then we would expect *lower* frequency lemmas to exhibit more defectiveness because there is already so little information both in the target form and in other cells of the paradigm.

For example, the lemma SKRAPA ([skraːpa]$_{[\text{NOM, SG}]}$ ‘scraper; stripper’) is very infrequent; it is only attested twice in the corpus (in the NOM.PL both times: [skrøːpʏr]). The learner has likely never seen this word before, and so there is no memorised form that she can fall back on. Furthermore, she might be more reluctant to apply either generalised rule (a → a) or (a → na) (7), since both those rules are only 50% reliable.

Finally, the more specific phonologically defined class of *-ɪːpa*-type inputs (ɪːpa → ɪpna) (7) are *too* specific, and would therefore be uninformative. Learners do seem to use highly specific rules that can be generalised for small phonologically-defined sets of lemmas (Albright and Hayes 2003), and when these rules work really well for a particular set of lexical items, they are called “islands of reliability”. Thus if there really is a class of *-ɪːpa*-type changes, then a learner would be reluctant to permeate the island. Immutable lexical sets defined by a strong pattern and a general reluctance to redefine existing rules have been established in the surface-true Whole-Word Morphology model (Baronian and Kulinich 2012). The solution for a learner who is confronted with these complications in the generalisations they form may be then to attempt to apply one of the unreliable rules, or else produce nothing at all.

While this approach provides a possible theoretical motivation for the uncertainty seen in these data, there remains an open question: how does a learner decide what an input pair looks like? Here, I made the choice of providing the NOM.SG and the GEN.PL, but as seen in the SKRAPA example, it is possible that the only form a learner is exposed to is *not* the NOM.SG, but something else like the NOM.PL, which may have a much less similar stem like [skrøːpʏr]. Albright and Hayes (2003) acknowledge that the MGL is a source-oriented learner (where correspondences are learnt between an input and an output), and that there are reasons to consider a product-oriented model, since learners formulate generalisations over patterns distributed between output forms (Bybee 1995). For example, a learner may induce the product-oriented generalisation that long vowels do not occur in closed syllables in disyllabic words separately from a source-oriented generalisation (a → na).

6.2 Irreconcilable ties as a mechanism for defectiveness in Fw.N-Free lemmas

In this section, I will show how suffixes and their morphological features are evaluated against weighted constraints. The grammar is not capable of generating a single unique output for the GEN.PL, and so this tie between winning candidates results in defectiveness. This mechanism uses a modified version of Optimal Interleaving (Wolf 2005), which posits a competition-based grammar where phonotactics and morphotactics are enforced at the same time. The input is lexically specified, and so it may contain phonological material and morphosyntactic features (8). The constraint set consists of both phonological and morphological constraints, and output candidates encode form-meaning mappings in ***morphs*** (8a-8c). Morphs are pairs of phonological exponents and their morphological features. These morphs can compete morphologically (8a and 8b) or phonologically (8b and 8c).

(8) Input: /bʊk-[+pl]/

a. ⟨[bʊk], BOOK⟩-⟨[-∅], [-pl]⟩

b. ⟨[bʊk], BOOK⟩-⟨[-s], [+pl]⟩

c. ⟨[bʊk], BOOK⟩-⟨[-z], [+pl]⟩

Sometimes, the morphological features in an output morph is more underspecified than the morphological features that are required in the input. Underspecification allows phonologically identical paradigm cells to be subsumed under a 1:1 mapping between a single phonological exponent and a single morphological feature set. This provides a principled way to distinguish between syncretism and homophony, where the latter consists of identical phonological exponents that map to different morphological feature sets. I propose that the homonymy observed within Icelandic nominal paradigms can all be analysed as truly syncretic forms, including the NOM.SG and GEN.PL cells in Fw.N-FREE paradigms. Underspecification is itself an analysis, and so I will first walk through how to carve up the paradigm in a way that minimises the morphological features that are necessary to compose a lexical entry.

Privative features (NOM, ACC, DAT, GEN, SG, PL) will not be sufficient, since there is not enough information in these features to capture the uniformity seen in these paradigms. I will therefore borrow binary features from Bierwisch's (1968) analysis of German syncretism. This breaks down the privative features into three binary features: [+/-governed], [+/-oblique], and [+/-plural]. The first two features specify case, and the last feature specifies number. The breakdown is shown below in tables 23 and 24.

Table 23: Syncretism in the paradigms for Fw LILJA 'lily', Mw ENDI 'end', Nw LUNGA 'lung'

<table>
<tr><th>Fw</th><th>SG [-pl]</th><th>PL [+pl]</th><th>Mw</th><th>SG</th><th>PL</th><th>Nw</th><th>SG</th><th>PL</th></tr>
<tr><td>NOM [-obl, -gov]</td><td>lilj-a</td><td rowspan="2">lilj-ur</td><td>NOM</td><td>end-i</td><td>end-ar</td><td>NOM</td><td rowspan="4">lung-a</td><td rowspan="2">lung-u</td></tr>
<tr><td>ACC [-obl, +gov]</td><td rowspan="3">lilj-u</td><td>ACC</td><td rowspan="3">end-a</td><td></td><td>ACC</td></tr>
<tr><td>DAT [+obl, +gov]</td><td>lilj-um</td><td>DAT</td><td>end-um</td><td>DAT</td><td>lung-um</td></tr>
<tr><td>GEN [+obl, -gov]</td><td>?lilj-a</td><td>GEN</td><td></td><td>GEN</td><td>lung-na</td></tr>
</table>

Next, I will compare the features of paradigm cells with identical phonological forms. The intersection of these feature sets will be retained as the syncretically defined feature set.

For example, syncretism between the NOM.PL and ACC.PL can be captured as [-o, -pl]. If there are no features at the intersection of two feature sets, then an empty set is returned instead, as is the case in the Mw paradigm seen below where *-a* has no feature specification.

Table 24: Underspecified features in the paradigms for Fw LILJA 'lily', Mw ENDI 'end', Nw LUNGA 'lung'

Fw morphs	⟨[+o, +g, +pl], -um⟩ ⟨[-o, +pl], -ur⟩ ⟨[-pl], -u⟩ ⟨[-g], -a⟩

Mw morphs	⟨[+o, +g, +pl], -um⟩ ⟨[-o, -g, +pl], -ar⟩ ⟨[-o, -g, -pl], -i⟩ ⟨[], -a⟩

Nw morphs	⟨[+o, +g, +pl], -um⟩ ⟨[-o, +pl], -u⟩ ⟨[+o, -g, -pl], -na⟩ ⟨[-pl], -a⟩

An interesting pattern of syncretism emerges that is unique to the Fw paradigm. The singular non-NOM and the GEN.PL are equally underspecified. This equal underspecification is the nature of defectiveness in the Fw paradigm. Whereas the singular non-NOM is underspecified for case, the GEN.PL is underspecified for number.

Next, the learner must be able to choose between the available suffixes. This can be accomplished through a series of morphological faithfulness constraints. These constraints (Wolf 2005:70–71) are as follows:

- MAX-M(F): For every instance ϕ of the feature F at the morpheme level, assign a violation-mark if there is not an instance ϕ' of F at the morph level, such that $\phi\mathfrak{R}\phi'$.

- DEP-M(F): For every instance ϕ' of the feature F at the morph level, assign a violation-mark if there is not an instance ϕ of F at the morpheme level, such that $\phi\mathfrak{R}\phi'$.

There are always going to be at least four potential output morphs (one for every partition in the Fw paradigm). This analysis employs Maximum Entropy (Hayes and Wilson 2008), which is a version of Harmonic Grammar (Smolensky 1986) where weighted constraints are directly mapped onto probabilities. This is first done by raising e to the summed harmony value of a candidate, and then the probability is calculated by dividing the exponentiated harmony value by all the exponentiated harmony values in a computation. The probabilities below can be read as "the probability that a lexical entry will be chosen amongst the four possible lexical entries".

Now we must establish a basic grammatical principle to guide the decision-making process. The Subset Principle (Halle 1997) states that a surface morpheme must contain a subset of the features that are specified in the input, and that if possible, the morpheme selected must match as many of the features of the surface morpheme as possible. To demonstrate this, I will show why *-um* wins instead of *-ur* in the DAT.PL cell of Fw paradigms (9). The winning candidate in this tableau (9) is fully specified, which entails that it will not violate any morphological faithfulness constraints. This means that this candidate will always win when it is being evaluated against the set of morphological constraints, since every other candidate violates at least two morphological constraints at a time.

(9) DAT.PL → *-um* [-ʏm]

/lɪlj/-[+o, +g, +pl]$_{\text{DAT.PL}}$	Dep(-o) 100	Max(+o) 70	Max(+g) 60	H	e^H	P
1. [lɪlj]-⟨[-pl], [-ʏ]⟩		-1	-1	-130	3.48111E-57	0%
2. [lɪlj]-⟨[-g], [-a]⟩		-1	-1	-130	3.48111E-57	0%
3. [lɪlj]-⟨[-o, +pl], [-ʏr]⟩	-1	-1		-170	1.295E-100	0%
4. ☞ [lɪlj]-⟨[+o, +g, +pl], [-ʏm]⟩				0	1	100%

Wolf (2005) argues that this typology emerges from a strict ranking of Dep-M over Max-M. I capture this relation between constraints in MaxEnt by positing that Dep-M constraints are more strongly weighted than Max-M constraints. Arbitrarily, I will set all Dep-M constraints at 100. The Max-M constraints are weighted at 60-95; the weights are largely arbitrary, although the ordering is meant to reflect the patterns of syncretism seen across inflection classes. Going from most common (and thus more highly weighted) to least common: number ≫ oblique ≫ governed. The exact weightings are a puzzle that will not be solved in this paper, but the relative weightings of Max-M(+pl) and Max-M(-gov) will be discussed later on in (12).

In this system, syncretism emerges from evaluations for different cells that output the same winners. These tableaux are shown below for the syncretism found in the non-NOM plural cells (10) as well as the NOM.PL and ACC.PL. In both the ACC and DAT tableaux, the Subset Principle is enforced by the strongly ranked Dep-M constraints, which work together with Max-M constraints to rule out every losing candidate. However, in the GEN.PL tableau, we see that candidate 2. is not picked out by the subset principle. Here, we observe that Max-M(-pl) must crucially be more strongly weighted than Max-M(-gov) in order to rule out the *-a* suffix.

(10) a. ACC.SG → *-u* [-ʏ]

/lɪlj/-[-o, +g, -pl]$_{\text{ACC.SG}}$	Dep(+pl) 100	Dep(+obl) 100	Dep(-gov) 100	Max(-pl) 80	Max(-obl) 70	Max(+gov) 60	H	e^H	P
1. ☞ [lɪlj]-⟨[-pl], [-ʏ]⟩					-1	-1	-130	3.48111E-57	100%
2. [lɪlj]-⟨[-g], [-a]⟩			-1	-1	-1	-1	-310	2.3373E-135	0%
3. [lɪlj]-⟨[-o, +pl], [-ʏr]⟩	-1			-1		-1	-240	5.8793E-105	0%
4. [lɪlj]-⟨[+o, +g, +pl], [-ʏm]⟩	-1	-1		-1	-1		-350	9.9296E-153	0%

b. DAT.SG → *-u* [-ʏ]

/lɪlj/-[+o, +g, -pl]$_{\text{DAT.SG}}$	Dep(+pl) 100	Dep(-obl) 100	Dep(-gov) 100	Max(-pl) 80	Max(+obl) 70	Max(+gov) 60	H	e^H	P
1. ☞ [lɪlj]-⟨[-pl], [-ʏ]⟩					-1	-1	-130	3.48111E-57	100%
2. [lɪlj]-⟨[-g], [-a]⟩			-1	-1	-1	-1	-310	2.3373E-135	0%
3. [lɪlj]-⟨[-o, +pl], [-ʏr]⟩	-1	-1		-1	-1	-1	-410	8.6949E-179	0%
4. [lɪlj]-⟨[+o, +g, +pl], [-ʏm]⟩	-1			-1			-180	6.71418E-79	0%

c. GEN.SG → *-u* [-Y]

/lılj/-[+o, -g, -pl]$_{GEN.SG}$	DEP(+pl) 100	DEP(-pl) 100	DEP(-obl) 100	DEP(+gov) 100	MAX(-pl) 80	MAX(+obl) 70	MAX(-gov) 60	H	e^H	P
1. ☞ [lılj]-⟨[-pl], [-Y]⟩						-1	-1	-130	3.48111E-57	100%
2. [lılj]-⟨[-g], [-a]⟩					-1	-1		-150	7.1751E-66	0%
3. [lılj]-⟨[-o, +pl], [-Yr]⟩	-1	-1	-1		-1	-1	-1	-510	3.2346E-222	0%
4. [lılj]-⟨[+o, +g, +pl], [-Ym]⟩	-1	-1		-1	-1		-1	-440	8.1363E-192	0%

In the NOM.PL and ACC.PL, we must introduce a new constraint MAX-M[-obl](-gov), which requires that an input feature [-gov] must be present in the output morph if the input feature set additionally contains [-obl]. This is because the constraints MAX-M(-pl) and MAX-M(-gov) are in conflict between the NOM.SG and NOM.PL paradigms, where the latter strongly favours a stronger weighted MAX-M(-pl) constraint, and the former prefers a stronger MAX-M(-gov) constraint.

(11) a. NOM.PL → *-ur* [-Yr]

/lılj/-[-o, -g, +pl]$_{NOM.PL}$	DEP(-pl) 100	DEP(+obl) 100	DEP(+gov) 100	DEP(-gov) 100	MAX[-obl](-gov) 95	MAX(+pl) 90	MAX(-obl) 70	MAX(+gov) 60	MAX(-gov) 60	H	e^H	P
1. [lılj]-⟨[-pl], [-Y]⟩	-1				-1	-1	-1		-1	-415	5.8585E-181	0%
2. [lılj]-⟨[-g], [-a]⟩						-1	-1			-160	3.25749E-70	1%
3. ☞ [lılj]-⟨[-o, +pl], [-Yr]⟩					-1				-1	-155	4.83454E-68	99%
4. [lılj]-⟨[+o, +g, +pl], [-Ym]⟩		-1	-1		-1		-1		-1	-425	2.6598E-185	0%

b. ACC.PL → *-ur* [-Yr]

/lılj/-[-o, +g, +pl]$_{ACC.PL}$	DEP(-pl) 100	DEP(+obl) 100	DEP(+gov) 100	DEP(-gov) 100	MAX[-obl](-gov) 95	MAX(+pl) 90	MAX(-obl) 70	MAX(+gov) 60	MAX(-gov) 60	H	e^H	P
1. [lılj]-⟨[-pl], [-Y]⟩	-1				-1	-1	-1	-1		-320	1.0611E-139	0%
2. [lılj]-⟨[-g], [-a]⟩				-1		-1	-1	-1		-320	1.0611E-139	0%
3. ☞ [lılj]-⟨[-o, +pl], [-Yr]⟩								-1		-60	8.75651E-27	100%
4. [lılj]-⟨[+o, +g, +pl], [-Ym]⟩		-1					-1			-170	1.4789E-74	0%

The problematic form of syncretism is the one that holds between the NOM.SG and GEN.PL. I contend that this occurs because the feature sets of the two cells are disjunctive both in case and number. This creates a competition for syncretism either within the plural or within the genitive. The competition alone does not result in defectiveness. Instead, defectiveness occurs when the constraint weightings are capable of selecting a unique winner in every Fw paradigm cell *except* the GEN.PL. For the purposes of this paper, these constraint weights are largely arbitrary, and the scope of the analysis will be limited to the Fw inflection class.

The goal here is to show that the weights must conspire to result in a tie between forms in the GEN.PL. In (12b), this tie occurs when candidates 2 and 4 have the same harmony scores; candidate 2 violates both MAX-M(+pl) and MAX-M(+obl), while candidate 4 violates DEP-M(+gov) and MAX-M(-gov).

(12) a. NOM.SG → *-a* [-a]

/lɪlj/-[-o, -g, -pl]$_{NOM.PL}$	Dep(+pl)	Dep(-pl)	Dep(+obl)	Dep(-obl)	Dep(+gov)	Max[-obl](-gov)	Max(+pl)	Max(-pl)	Max(+obl)	Max(-obl)	Max(-gov)	H	e^H	P
	100	100	100	100	100	95	90	70	70	60	60			
1. [lɪlj]-⟨[-pl], [-ʏ]⟩						-1				-1	-1	-225	1.92195E-98	0%
2. ☞ [lɪlj]-⟨[-g], [-a]⟩								-1		-1		-150	7.1751E-66	100%
3. [lɪlj]-⟨[-o, +pl], [-ʏr]⟩	-1					-1		-1			-1	-335	3.246E-146	0%
4. [lɪlj]-⟨[+o, +g, +pl], [-ʏm]⟩	-1		-1		-1	-1		-1		-1	-1	-605	1.7858E-263	0%

b. GEN.PL → *-a* [-a]

/lɪlj/-[+o, -g, +pl]$_{ACC.PL}$	Dep(+pl)	Dep(-pl)	Dep(+obl)	Dep(-obl)	Dep(+gov)	Max[-obl](-gov)	Max(+pl)	Max(-pl)	Max(+obl)	Max(-obl)	Max(-gov)	H	e^H	P
	100	100	100	100	100	95	90	70	70	60	60			
1. [lɪlj]-⟨[-pl], [-ʏ]⟩		-1					-1		-1		-1	-320	1.0611E-139	0%
2. ☹ ☞ [lɪlj]-⟨[-g], [-a]⟩							-1		-1			-160	3.25749E-70	50%
3. [lɪlj]-⟨[-o, +pl], [-ʏr]⟩				-1					-1		-1	-230	1.295E-100	0%
4. ☞ [lɪlj]-⟨[+o, +g, +pl], [-ʏm]⟩					-1						-1	-160	3.25749E-70	50%

The tie in (12b) is uniquely problematic for two reasons. Firstly, candidate 2 is underspecified for number. In all the tableaux from (10-12), tableau (12b) is the only one in which a the desired winner violates Max-M(+pl). Although candidate 2 is also underspecified for [oblique], underspecification of case features is not rare: both the desired winners in the DAT.SG and the GEN.SG (10) violate Max-M(+obl). Secondly, the undesirable winner (candidate 4) is the only winner in all these tableaux that would require the learner to violate the Subset Principle. It should be noted again that these weights are arbitrary; functionally, what matters is that defectiveness occurs from *any* tie that is unique to GEN.PL morpheme selection. Therefore, it is possible that a tie could arise with candidate 3, for instance, depending on the initial state of the learner's constraint weights[18], but it would still result in a violation of Dep-M(-obl), which in turn conflict with the Subset Principle.

This approach reframes Fw.N-Free defectiveness as a learning problem. A candidate tie is thus intended to convey the uncertainty of the learner in the process of lexical retrieval, and not the optional availability of two affixes. To that end, there are no *-um* suffixed GEN.PL forms, nor does this model intend to predict that there should be. When a learner encounters a tie between input-output mappings, I propose that the learner opts to fail in response to this uncertainty. In section 6.1, I presented Fw.Other uncertainty as the phonological conditioning of two equally possible surface forms. In this section, I propose that Fw.N-Free uncertainty blocks surface mapping through morphologically driven indeterminacy.

It is not clear whether higher frequency lemmas should be expected to exhibit more defectiveness than lower frequency ones. On the one hand, this speculation may be incorrect, and higher frequency lemmas are actually less defective. If this is the case, then as lemma input frequency goes up, we would expect candidates' probabilities to scale up since the learner would become more confident in one or more constraints' weights. On the other hand, this speculation may be correct. It could be the case that as input frequency goes up, a learner encounters more conflicting evidence (e.g., half the data are defective and half are not), which proliferates the tie. Since I have only yet speculated about the behaviour of higher frequency lemmas, I will leave this up to further empirical tests in the future.

[18] It could also be possible to observe a tie with candidate 1, but this would require Dep-M(-pl) to weigh less than a Max-M constraint.

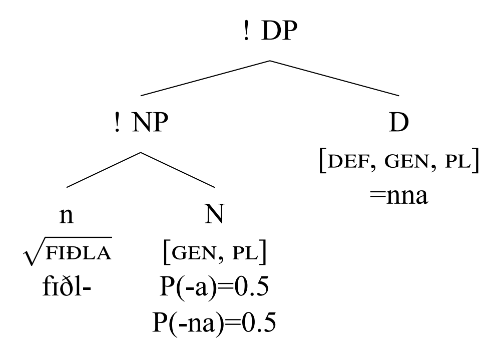
6.3 Definite constructions depend on the defectiveness of the base

The models in section 5 demonstrated three important effects with respect to definiteness: Fw lemmas are more underattested than Mw lemmas, definite tokens were overall less frequent than indefinite tokens, and Fw lemmas were more underattested by definite tokens than Mw lemmas were. This suggests that whichever of the two defectiveness types are inhibiting word formation (see sections 6.1 and 6.2), the defectiveness is likely inherited by a word's definite form. This inheritance (or dependence) can be represented in a syntactic structure, where word formation fails when a root and a suffix fail to properly merge (or earlier), which is marked with an exclamation point at either the head or phrase level of the NP.

I created the syntactic structures in figures 7 and 8 using an extremely condensed version of the word-internal structure in Harðarson (2016). The root, marked as a little n, is assumed to be acategorical. Noun category is established by merging n with constituents higher up in the internal word structure, which I have collapsed into a single NP. The definite marker is assumed to be a clitic for phono-morphological reasons (e.g., variable ordering within a word; does not trigger vocalic alternations in the stem). When this defective NP specifies a DP, the D-head is able to agree with the morpho-syntactic features of the lower N-head (regardless of whether a phonological form can be retrieved), but the defectiveness of the NP ultimately percolates up to the DP level.

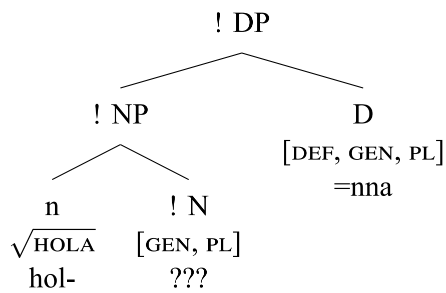
In figure 7, defectiveness begins at the level of the NP in gen.pl forms of some Fw.Other lemmas. The N-head itself is not defective since lexical retrieval is possible, but certain phonological environments may result in the availability of two equally probable allomorphs. This indeterminacy causes a failure to merge, i.e., inflect, which in turn passes up defectiveness to the DP phrase level.

Figure 7: Failed derivation of a defective base (Fw.Other)



In figure 8, there is a two-pronged approach that leads to defectiveness in the gen.pl forms of some Fw.N-Free lemmas: a failure at the morphology and a failure at the phonology. Given a set of features [gen, pl] for the inflectional suffix, the derivation crashes when the lexicon is accessed. Since the appropriate morph cannot be retrieved, there is no morphological or phonological information that a learner can employ for word formation. This means that the acategorical root remains uninflected, and the NP has a faulty head. A bare root cannot occur in spec-DP on its own, and so the whole derivation fails.

Figure 8: Failed derivation of a defective base (Fw.N-Free)



It is not clear whether there are any lemmas whose gen.pl forms surface in the definite but *not* the indefinite. Since this study has only looked at class-based defectiveness, future work at the lemma level would help to answer this question.

7 Conclusion

Although this study has presented multiple methodological challenges, it begins to scratch the surface of determining the strongest correlates of defectiveness across languages. With respect to the research questions posed in section 3, the intuitions of native Icelandic speakers have indeed been confirmed: Fw gen.pl nouns are certainly underrepresented in this data. The motivations of this underrepresentation seem to be linked to interparadigmatic syncretism avoidance, but there is still more work to be done, with respect to corpus curation, for these motivations to be generalisable. With that being said, this work not only provides large-scale empirical support for native speaker intuitions, but also has provided a frequentist methodology that can be used to test potential causes of defectiveness by mapping these causes to predictor variables.

Future work on this particular corpus would involve a larger sample size, more attention to phonological classes that are independent from morphosyntactic classification, and a more in-depth exploration of variation. There is a growing body of literature on empirical methods for modelling defectiveness. The Albright (2003) corpus study on Spanish gaps found that infrequent or unfamiliar lexemes were more likely to be gapped, and irregular inflections were also more likely to be gapped even amongst high frequency tokens. Daland et al. (2007) used a Bayesian learner with varying degrees of analogical pressures to model token frequencies of defective verbs in a Russian corpus. The authors found that gaps persisted in two conditions: when the learner was given a lexicon with no additional grammatical information, and when there were few analogical forms of the same inflection class that were very similar. For instance, if an Icelandic learner is exposed to few genitive plural wordforms, and those wordforms have similar phonological shapes, then Daland et al. (2007) predict that the pressure of competition would more strongly prevent the learner from filling the gap than if the wordforms were less similar or if there was simply more evidence.

Finally, one potential source of defectiveness may stem from pragmatic or communicative goals that may be encoded in the register or medium where a token's usage is highly restricted. Löwenadler (2010) analysis of Swedish gaps delves into the pragmatic motivation for defectiveness. The author proposes that while all native speakers possess "morphological creativity", which is the ability to use word formation strategies (WFS) to coin new words, the notion of "morphological productivity" only relegates a specific subset of creative WFS if they satisfy a communicative need. Löwenadler argues that more infrequent and irregular forms are less productive, and so by extension, require a stronger communicative need or else a gap will proliferate. This places paradigm gaps in a state of suspended animation, whereby wordforms that require less productive morphology will not inflect unless external pressures are strong enough to call for their use. This may also translate into a distinction between productive defectiveness (e.g., “I will not produce that form”) and perceptive defectiveness (e.g., “I will not accept that form”).

Glossary

1	first person	LOC	locative
3	third person	M	masculine
ACC	accusative	NOM	nominative
DAT	dative	PL	plural
DEF	definite	PST	past
GEN	genitive	SG	singular
INDF	indefinite		

References

Ackerman, F., Blevins, J., and Malouf, R. (2009). Parts and wholes: Implicative patterns in inflectional paradigms. In *Analogy in Grammar*, pages 54–82. Journal Abbreviation: Analogy in Grammar.

Albright, A. (2003). A quantitative study of Spanish paradigm gaps. pages 1–14, Somerville, MA. Cascadilla Press.

Albright, A. and Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition*, 90(2).

Ayala, A. and Hansson, G. (2021). Gradient defectiveness in icelandic noun inflection: a quantitative study. Talk presented at the 5th Annual International Morphology Meeting on 29 August 2021 at Ohio University.

Baerman, M. (2011). Defectiveness and homophony avoidance. *Journal of Linguistics*, 47(1):1–29.

Baronian, L. and Kulinich, E. (2012). Paradigm gaps in Whole Word Morphology *. In *Irregularity in Morphology (and beyond)*, pages 81–100. Journal Abbreviation: Irregularity in Morphology (and beyond).

Bjarnadóttir, K. (2021). Beygingarlýsing íslensks nútímamáls. [the database of modern icelandic inflection.].

Blevins, J. P. (2016). Revival of the WP model. In *Word and Paradigm Morphology*. Oxford University Press, Oxford.

Blevins, J. P., Ackerman, F., and Malouf, R. (2018). Word and Paradigm Morphology. In Audring, J. and Masini, F., editors, *The Oxford Handbook of Morphological Theory*, pages 264–284. Oxford University Press.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, 10(5):425–455.

Cleasby, R. and Vigfusson, G. (1874). *An Icelandic-English Dictionary*. Oxford University Press, London, UK.

Daland, R., Sims, A., and Pierrehumbert, J. (2007). Much ado about nothing: A social network model of Russian paradigmatic gaps. In *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 936–943, Prague, Czech Republic. Association for Computational Linguistics.

Eberhard, D. M., Simons, G. F., and Fennig, C. D. (2021). *Ethnologue: Languages of the World*. SIL International, Dallas, TX, 24 edition.

Finkel, R. and Stump, G. (2007). Principal parts and morphological typology. *Morphology*, 17:39–75.

Halle, M. (1997). Distributed Morphology: Impoverishment and Fission. *MIT Working Papers in Linguistics*, 30:425–449.

Halle, M. and Marantz, A. (1993). *Distributed morphology and the pieces of inflection*, pages 111–176. The MIT Press.

Hansson, G. (2015). *Oxford Handbook of Inflection*, chapter Phonology, pages 161–195.

Harðarson, G. R. (2016). Word building and the Icelandic noun phrase. In *Proceedings of the 39th Annual Penn Linguistics Conference*, pages 323–332. University of Pennsylvania Working papers in Linguistics.

Hayes, B. and Wilson, C. (2008). A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry*, 39(3):379–440.

Ingason, A. K. (2016). *Realizing Morphemes in the Icelandic Noun Phrase*. PhD thesis, University of Pennsylvania.

Jónsson, S. (1941/1927). *A Primer of Modern Icelandic*. Oxford University Press, London, UK.

Löwenadler, J. (2010). Restrictions on productivity: defectiveness in swedish adjective paradigms. *Morphology*, (20):71–107.

Müller, G. (2005). Syncretism and iconicity in icelandic noun declensions: a distributed morphology approach.

Pertsova, K. (2016). Transderivational relations and paradigm gaps in russian verbs. *Glossa: a journal of general linguistics*, 1(1):1–34.

Sims, A. D. (2015). *Inflectional Defectiveness*, volume 148. Cambridge University Press, Cambridge.

Smolensky, P. (1986). Information processing in dynamical systems: foundations of Harmony Theory. In *Parallel Distributed Processing*, pages 194–281. MIT Press, Cambridge.

Steingrímsson, S., Helgadóttir, S., Rögnvaldsson, E., Barkarson, S., and Guðnason, J. (2018). Risamálheild: A very large Icelandic text corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan.

Steriade, D. (1998). Lexical conservatism and the notion base of affixation. *Linguistics in the Morning Calm, Selected Papers from SICOL 1997, Linguistic Society of Korea*, pages 157–179.

Stump, G. (2001). *Inflectional Morphology: A Theory of Paradigm Structure*.

Wolf, M. A. (2005). *Optimal interleaving: Serial phonology -morphology interaction in a constraint-based model*. Ph.D., University of Massachusetts Amherst, United States – Massachusetts. ISBN: 9780549915737.

8 Appendix

The tables below were compiled by Gunnar Ó. Hansson and are to be read as follows: 150 tokens were chosen at random from each lemma frequency bin, and these tokens were returned in their sentential contexts and hand-checked. Since these tokens were randomly sampled, only tokens from frequency bin 10^0 have no chance of occurring twice, while there was much more repetition at higher frequencies (some with similar contexts, since it appeared that the same source would contain multiple instances of the same token).

lemma frequency bin	number of tokens	# of false positives / 150	percentage
10^0	136	65/136	47.8%
$<10^1$	391	48/150	32.0%
$<10^2$	482	22/150	14.7%
$<10^3$	498	10/150	6.7%
$<10^4$	499	2/150	1.3%
$<10^5$	500	0/150	0.0%

Table 25: Error-checking results done by GÓH: false positive rates for weak feminine nouns in the INDF.NOM.PL and INDF.ACC.PL.

lemma frequency bin	number of tokens	# of false positives / 150	percentage
10^0	81	66/81	81.5%
$<10^1$	369	96/150	64.0%
$<10^2$	474	85/150	56.7%
$<10^3$	491	39/150	26.0%
$<10^4$	496	31/150	20.7%
$<10^5$	500	19/150	12.7%

Table 26: Error-checking results done by GÓH: false positive rates for weak masculine nouns in the INDF.NOM.PL and INDF.ACC.PL.