

Measuring phonological change in Old Norse manuscripts

Jade J. Sandstedt

jsandstedt@gmail.com

jsandstedt.hcommons.org

Humboldt University of Berlin

29. May 2019

Outline

Topic:

- ▶ Philological challenges in historical phonological research

Problem:

- ▶ competing orthographic vs. phonological vs. etymological factors in spelling variation

Solution:

- ▶ rich linguistic annotations

Outline

- 1 Background
 - Old Norwegian vowel harmony
 - Old Norwegian vowel harmony decay
- 2 Old Norwegian corpus and grapho-phonological methods
- 3 Sanity check
 - Lexical frequencies
 - Syllable and vowel frequencies
- 4 Tracking vowel harmony decay
 - Visualising decay
- 5 Conclusions

Background: Old Norwegian vowel harmony

Old Norwegian (c 1200–1350) displays a form of vowel height harmony (I)

- resulting in [-i]/[-e] and [-u]/[-o] suffixal alternations

Background: Old Norwegian vowel harmony

Old Norwegian (c 1200–1350) displays a form of vowel height harmony (I)

- resulting in [-i]/[-e] and [-u]/[-o] suffixal alternations

(I) Height harmony in Old Norwegian (Sandstedt 2017, 2018)

HIGH	h <u>ú</u> s-i	<hufi>	h <u>ú</u> s-um	<hufū>	‘house’-DAT.SG./PL.
	sk <u>i</u> p-i	<fkɪpɪ>	sk <u>i</u> p-um	<fkɪpum>	‘ship’-DAT.SG./PL.
NON-HIGH	lj <u>ó</u> s-e	<lɪofe>	lj <u>ó</u> s-om	<lɪofom>	‘light’-DAT.SG./PL.
	s <u>e</u> gl-e	<fegle>	s <u>e</u> gl-om	<feglō>	‘sail’-DAT.SG./PL.

Background: Old Norwegian vowel harmony

Old Norwegian (c 1200–1350) displays a form of vowel height harmony (I)

- resulting in [-i]/[-e] and [-u]/[-o] suffixal alternations

(I) **Height harmony in Old Norwegian** (Sandstedt 2017, 2018)

HIGH	h <u>ú</u> s-i	<hufi>	h <u>ú</u> s-um	<hufū>	‘house’-DAT.SG./PL.
	sk <u>i</u> p-i	<fkɪpɪ>	sk <u>i</u> p-um	<fkɪpum>	‘ship’-DAT.SG./PL.
NON-HIGH	lj <u>ó</u> s-e	<lɔfe>	lj <u>ó</u> s-om	<lɔfom>	‘light’-DAT.SG./PL.
	se <u>g</u> l-e	<fegle>	se <u>g</u> l-om	<feglō>	‘sail’-DAT.SG./PL.

- cf. non-harmonic Old Icelandic *ljós-i* and *segl-i*

Vowel harmony: the broader perspective

Very generally defined, **vowel harmony** is a process in which vowels in a word show systematic correspondence for some feature.

Old Norwegian (Germanic) height harmony

[+high] hús-i ‘house’-DAT.SG.

[-high] ljós-e ‘light’-DAT.SG.

Finnish (Finno-Ugric) backness harmony (Ringen 1975)

[+back] pouta-na ‘fine weather’-ESS.

[-back] pöytnä-nä ‘table’-ESS.

Yoruba (Atlantic-Congo) tongue root harmony (Ọla Orié 2001, 2003)

[+ATR] ògèdè ‘incantations’

[-ATR] ɔ̀gèdè ‘banana, plantain’

Prevalence and motivations for harmony

Harmony systems are

Prevalence and motivations for harmony

Harmony systems are

- ▶ articulatorily and perceptually motivated,
 - ▶ eases articulation, makes sequences more predictable, enhances perceptually weak cues, etc. (Suomi 1983, Gallagher 2010, Walker 2005)

Prevalence and motivations for harmony

Harmony systems are

- ▶ articulatorily and perceptually motivated,
 - ▶ eases articulation, makes sequences more predictable, enhances perceptually weak cues, etc. (Suomi 1983, Gallagher 2010, Walker 2005)
- ▶ acquired early—at about the same time as mastery of the sound inventory
 - ▶ few to no harmony violations by ca. 2;6 years (MacWhinney 1978, Leiwo, Kulju & Aoyama 2002, Altan 2007)

Prevalence and motivations for harmony

Harmony systems are

- ▶ articulatorily and perceptually motivated,
 - ▶ eases articulation, makes sequences more predictable, enhances perceptually weak cues, etc. (Suomi 1983, Gallagher 2010, Walker 2005)
- ▶ acquired early—at about the same time as mastery of the sound inventory
 - ▶ few to no harmony violations by ca. 2;6 years (MacWhinney 1978, Leiwo, Kulju & Aoyama 2002, Altan 2007)
- ▶ cross-linguistically very common and diachronically robust,
 - ▶ e.g. millenia old backness harmony in Altaic languages (Harrison, Dras & Kapicioglu 2006)

Prevalence and motivations for harmony

Harmony systems are

- ▶ articulatorily and perceptually motivated,
 - ▶ eases articulation, makes sequences more predictable, enhances perceptually weak cues, etc. (Suomi 1983, Gallagher 2010, Walker 2005)
- ▶ acquired early—at about the same time as mastery of the sound inventory
 - ▶ few to no harmony violations by ca. 2;6 years (MacWhinney 1978, Leiwo, Kulju & Aoyama 2002, Altan 2007)
- ▶ cross-linguistically very common and diachronically robust,
 - ▶ e.g. millenia old backness harmony in Altaic languages (Harrison, Dras & Kapicioglu 2006)

Question

If harmony is so natural and beneficial, **why don't all languages display harmony?**

Harmony doesn't last forever

Despite stability of harmony systems,

- ▶ harmony systems do decay
 - e.g. Turkish vs. Uzbek (Turkic; Csató & Johanson 1998; Sjöberg 1963).

Harmony decay in Turkic languages

(2) Turkic backness harmony lost in Uzbek

BACK	d <u>o</u> st-lar		‘friend’-PL.	d <u>o</u> ‘st-lar		‘friend’-PL.
	k <u>u</u> l-lar		‘slave’-PL.	q <u>u</u> l-lar		‘slave’-PL.
FRONT	<u>e</u> t-ler	* <u>e</u> t-lar	‘meat’-PL.	<u>e</u> t-lar	* <u>e</u> t-ler	‘meat’-PL.
	di <u>ş</u> -ler	*di <u>ş</u> -lar	‘tooth’-PL.	t <u>i</u> sh-lar	*t <u>i</u> sh-ler	‘tooth’-PL.

(a) Turkish – [-lar] / [-ler]

(b) Uzbek – [-lar]

Harmony decay in Nordic languages

(3) Early Old Norse height harmony lost in Icelandic

HIGH	h <u>u</u> s-um		'house'-DAT.PL.	h <u>u</u> s-um		'house'-DAT.PL.
	sk <u>i</u> p-um		'ship'-DAT.PL.	sk <u>i</u> p-um		'ship'-DAT.PL.
NON-/	lj <u>o</u> s-om	*lj <u>o</u> s-um	'light'-DAT.PL.	lj <u>o</u> s-um	*lj <u>o</u> s-om	'light'-DAT.PL.
HIGH	s <u>e</u> gl-om	*s <u>e</u> gl-um	'sail'-DAT.PL.	s <u>e</u> gl-um	*s <u>e</u> gl-om	'sail'-DAT.PL.

(a) Old Norwegian – [-um] / [-om]

(b) (Old) Icelandic – [-um]

Harmony decay in Nordic languages

(3) Early Old Norse height harmony lost in Icelandic

HIGH	h <u>ú</u> s-um		'house'-DAT.PL.	h <u>ú</u> s-um		'house'-DAT.PL.
	sk <u>i</u> p-um		'ship'-DAT.PL.	sk <u>i</u> p-um		'ship'-DAT.PL.
NON-/	lj <u>ó</u> s-om	*lj <u>ó</u> s-um	'light'-DAT.PL.	lj <u>ó</u> s-um	*lj <u>ó</u> s-om	'light'-DAT.PL.
HIGH	se <u>g</u> l-om	*se <u>g</u> l-um	'sail'-DAT.PL.	se <u>g</u> l-um	*se <u>g</u> l-om	'sail'-DAT.PL.

(a) Old Norwegian – [-um] / [-om]

(b) (Old) Icelandic – [-um]

Question

If harmony is so natural and historically stable, **what motivates the loss of harmony?**

Sources of evidence

Sources of evidence

- ▶ comparisons between harmonic/non-harmonic dialects ([Kavitskaya 2013](#))

Sources of evidence

- ▶ comparisons between harmonic/non-harmonic dialects ([Kavitskaya 2013](#))
- ▶ diachronic comparisons before and following harmony decay ([McCollum 2015](#), [Bobaljik 2018](#))

Sources of evidence

- ▶ comparisons between harmonic/non-harmonic dialects (Kavitskaya 2013)
- ▶ diachronic comparisons before and following harmony decay (McCollum 2015, Bobaljik 2018)
- ▶ agent-based computational modelling of potential trajectories of vowel harmony evolution/dissolution (Harrison, Dras & Kapicioglu 2006; Mailhot 2010)

Suspected causes of decay

Suspected causes of decay

- ▶ changes in vowel inventories (mergers/splits),

Suspected causes of decay

- ▶ changes in vowel inventories (mergers/splits),
- ▶ emergence of disharmonic morphemes,

Suspected causes of decay

- ▶ changes in vowel inventories (mergers/splits),
- ▶ emergence of disharmonic morphemes,
- ▶ language contact (i.e. via the influx of disharmonic foreign loanwords)

Suspected causes of decay

- ▶ changes in vowel inventories (mergers/splits),
- ▶ emergence of disharmonic morphemes,
- ▶ language contact (i.e. via the influx of disharmonic foreign loanwords)

Problem

We lack empirical evidence

- ➡ crucial missing link in the typological record: the transition from a harmonic to non-harmonic language.
- ➡ unclear how and why these factors might converge on the loss of harmony

Old Norse corpus: manuscripts

The Old Norwegian historical record can fill this typological lacuna

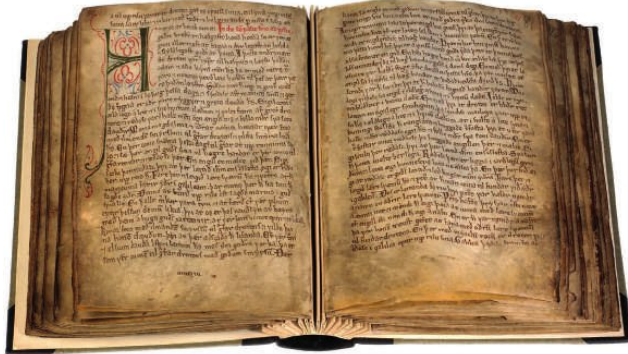


Figure 3: Old Norwegian Homily Book (AM 619 4to; c 1200)

Old Norse corpus: charters

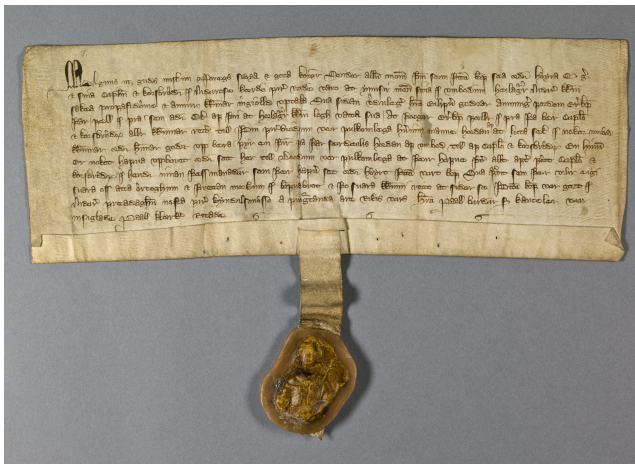


Figure 4: A charter from King Magnus VII – Niðaróss (Trondheim), 29. Jan. 1333

Old Norse corpus: runic inscriptions



Figure 5: Runestone from Kingittorsuaq (Greenland) with height harmony

Old Norwegian historical record

Old Norwegian philological material:

- ▶ provides a sizeable corpus of manuscripts, charters, and runic inscriptions
- ▶ covering pre-, transitional, and post-decay stages of the language

 making Old Norwegian a typologically highly significant specimen

Outline

- 1 Background
 - Old Norwegian vowel harmony
 - Old Norwegian vowel harmony decay
- 2 Old Norwegian corpus and grapho-phonological methods
- 3 Sanity check
 - Lexical frequencies
 - Syllable and vowel frequencies
- 4 Tracking vowel harmony decay
 - Visualising decay
- 5 Conclusions

Old Norwegian harmony and Norse philology

Old Norwegian harmony and Norse philology

Harmony is found in the earliest writing on parchment (c mid-12th century)

- ▶ decaying gradually over the course of the late 13th and 14th centuries (Flom 1934, Seip 1955, Hødnebo 1977, Hagland 1978)

Old Norwegian harmony and Norse philology

Old Norwegian harmony and Norse philology

Dette [vokalharmoni]systemet kan følges fra eldste skrifttid og et godt stykke inn i 1300-tallet som en slags norm. Henimot slutten av hundreåret inntreer en jevn tilbakegang med stadig flere unntak fra regelen.

This [vowel harmony] system can be seen from the oldest writings and up to a good ways into the 1300s as a kind of norm. Towards the end of the century, there is a steady decline with ever-increasing exceptions to the rule. (Hødnebo 1977, 379)

Methodological challenges

Hødnebø's generalisation is statistical

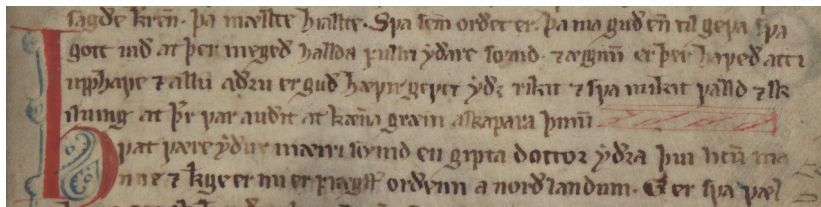
- ▶ we need to be able to quantify Old Norwegian harmony patterns

Problem: Old Norwegian corpus is largely inaccessible

- ▶ and poses significant philological problems

Harmony data collection

- (4) Old Norwegian vowel height harmony in prose (De la Gardie 8, fol. 16r21–26)



21 sagðe konongrenn. þa mællte hiallte. Sva sem orðet er. þa ma guð enn til geva sva
22 gott rað at þer meged hallda fullri yðare sæmd. oc æignum er þer haveð att i
23 upphave oc allu aðru er guð hævir gevet yðr rikit oc sva mikit valld oc sk.
24 ilning at þer var auðit at kænna græina skapara þinum.
25 Hvat vere yður mæiri sæmd en gipta dottor yðra þui licum ma.
26 nne oc kononge er nu er frægstr orðenn a norðlandum. Ger sva vel

Outline

- 1 Background
 - Old Norwegian vowel harmony
 - Old Norwegian vowel harmony decay
- 2 Old Norwegian corpus and grapho-phonological methods
- 3 Sanity check
 - Lexical frequencies
 - Syllable and vowel frequencies
- 4 Tracking vowel harmony decay
 - Visualising decay
- 5 Conclusions

Medieval Nordic Text Archive (MENOTA):

<https://menota.org/forside.xhtml>

- ▶ an increasing, digitised sample of Old Norwegian manuscript material
 - ▶ many of which are lemmatised and morphologically parsed

The elicited corpus

Abbr.	Signature	MS or work title	Date	Provenance	Words
AM243	AM 243 bæ fol	<i>King's Mirror</i>	c 1275	Bergen	63910
DG4_7_h1	De la Gardie 4-7, fols. 17va6-29v	<i>Strengleikar</i> -hand 1	c 1270	Bergen	19813
DG4_7_h2	De la Gardie 4-7, fols. 30r-43v	<i>Strengleikar</i> -hand 2	c 1270	Bergen	18640
DG8	De la Gardie 8 fol, fols. 70v-110v	<i>Legendary saga of St. Olaf</i>	c 1225-50	Trøndsk	41142
H6	Holm perg 6 fol	<i>Saga of Barlaam and Josaphat</i>	c 1275	Eastern	76411
H17	Holm perg 17 4to	<i>Saga of Archbishop Thómas</i>	c 1300	Uncertain	59884

Table 1: The selected Old Norwegian manuscript corpus

Structure of MENOTA transcriptions

- (5) MENOTA transcription of <hofðíngianom> ‘chieftain’-DAT.M.SG.-DEF.
(Holm perg 6 fol.)

```
<w xml:id='w034581' lemma='hofðíngi' me:msa='xNC gM nS cD sD'>  
<me:dipl>hofðíngianom</me:dipl>  
</w>
```

With the help of Pavel Iosad (Edinburgh), we have developed an algorithm, collecting:

- ▶ word IDs
- ▶ vowel patterns
- ▶ morphological annotations
- ▶ lemmas
- ▶ etc.

A vowel harmony database

Using the data from the MENOTA corpus,

- ▶ vowels are organised into pairwise sequences and evaluated for relative height agreement (2)

Harmonic span	V ₁	V ₂	V ₁ _high	V ₂ _high	VH
{hofðíng} ₁ ianom	<o>	<í>	False	True	False
hof {ðíngia} ₂ nom	<í>	<a>	True	False	False
hofðíng {ianom} ₃	<a>	<o>	False	False	True

Table 2: Division into pairwise harmonic spans

Controlling for variation

An orthographic database like (2) is useful, but it needs to recover orthographic variation

- * e.g. spelling variation for [ɔ]
 - ▶ e.g. <hofðingja> and <hafðingja> for normalised *hofðingja*

Grapho-phonology

We need a way to triangulate between distinct etymological, phonological, and orthographic values for each given segment

- ▶ e.g. <hofðingia> = <o-i> = [ɔ-i]
- ▶ e.g. <hafðingia> = <a-i> = [ɔ-i]

Phonological/etymological annotations

Etymological annotations based on Holthausen (1948)

- ▶ encoded for the 600 most common lexemes in the corpus (185,534 words)
 - ▶ dataset is available online at <http://dx.doi.org/10.17613/gj6n-js33>

Phonological/etymological annotations

Phonological/etymological annotations

Some scribes fail to represent all contrasts in writing

- ▶ though they are recoverable from phonological behaviour (6)
 - ▶ e.g. height harmonic /e/ vs. height disharmonic /æ/

Phonological/etymological annotations

Some scribes fail to represent all contrasts in writing

- ▶ though they are recoverable from phonological behaviour (6)
- ▶ e.g. height harmonic /e/ vs. height disharmonic /æ/

(6) (Non-)etymological *e/*æ spellings in Old Norwegian manuscripts

DG8 – c (1225–50)	[e , æ] – < e , æ >		
*e <ge <u>v</u> - e >	‘give’-3.SG.PRES.SUBJ.	<re <u>k</u> - et >	‘drive’-PRET.PART.
*æ <h <u>æ</u> v- i >	‘have’-3.SG.PRES.SUBJ.	<te <u>k</u> - it >	‘take’-PRET.PART.
H6 – c (1275)	[e , æ] – < e >		
*e <ge <u>v</u> - e >	‘give’-3.SG.PRES.SUBJ.	<re <u>k</u> - et >	‘drive’-2.PL.IMP.
*æ <he <u>v</u> - i >	‘have’-3.SG.PRES.SUBJ.	<te <u>k</u> - it >	‘take’-PRET.PART.

Phonological/etymological annotations

Some scribes fail to represent all contrasts in writing

- ▶ though they are recoverable from phonological behaviour (6)
- ▶ e.g. height harmonic /**e**/ vs. height disharmonic /**æ**/

(6) (Non-)etymological ***e**/***æ** spellings in Old Norwegian manuscripts

DG8 – c (1225–50)	[e , æ] – < e , æ >		
* e <ge v - e >	‘give’-3.SG.PRES.SUBJ.	<re k - e t>	‘drive’-PRET.PART.
* æ <h æ v- i >	‘have’-3.SG.PRES.SUBJ.	<ta k - i t>	‘take’-PRET.PART.
H6 – c (1275)	[e , æ] – < e >		
* e <ge v - e >	‘give’-3.SG.PRES.SUBJ.	<re k - e t>	‘drive’-2.PL.IMP.
* æ <he v - i >	‘have’-3.SG.PRES.SUBJ.	<te k - i t>	‘take’-PRET.PART.

👉 with etymological annotations we can distinguish the contrasts in (6) despite orthographic underrepresentation in H6

Phonological/etymological annotations

(7) Example annotation of root-initial vowels in *standa* 'stand' word forms

```
1  vh_df <- vh_df %>%
2    mutate (etym1 = replace (etym1, which (lemma == "standa"
3                                     & tense == "preterite"
4                                     & mood == "indicative"), "o:"
5                                     " ")) %>%
6    mutate (etym1 = replace (etym1, which (lemma == "standa"
7                                     & tense == "preterite"
8                                     & mood == "subjunctive"), "ø:"
9                                     " ")) %>%
10   mutate (etym1 = replace (etym1, which (lemma == "standa"
11                                     & tense == "present"
12                                     & number == "plural"
13                                     & mood == "indicative"
14                                     & person == "1. person"
15                                     & v1 %in% c("a", "o")), " "
16                                     " ")) %>%
17   mutate (etym1 = replace (etym1, which (lemma == "standa"
18                                     & tense == "present"
19                                     & number == "singular"
20                                     & mood == "indicative"
21                                     & v1 %in% c("e", "æ", "ǣ")),
22                                     " " ))
```

Phonological/etymological annotations

(8) Etymological vs. graphic representations of the *standa* 'stand' finite verb paradigm

Present indicative				Present subjunctive			
Singular		Plural		Singular		Plural	
1.	<i>*stend</i> <stænd>	<i>*stǣndum</i> <standum>		<i>*stande</i> <stande>		<i>*standem</i> <standem>	
2.	<i>*stendr</i> <stændr>	<i>*standet</i> <standet>		<i>*stander</i> <stander>		<i>*standet</i> <standet>	
3.	<i>*stendr</i> <stændr>	<i>*standa</i> <standa>		<i>*stande</i> <stande>		<i>*stande</i> <stande>	

Preterite indicative				Preterite subjunctive			
Singular		Plural		Singular		Plural	
1.	<i>*sto:ð</i> <stoð>	<i>*sto:ðom</i> <stoðom>		<i>*stø:ðe</i> <stœðe>		<i>*stø:ðem</i> <stœðem>	
2.	<i>*sto:tt</i> <stott>	<i>*sto:ðoð</i> <stoðoð>		<i>*stø:ðer</i> <stœðer>		<i>*stø:ðeð</i> <stœðeð>	
3.	<i>*sto:ð</i> <stoð>	<i>*sto:ðo</i> <stoðo>		<i>*stø:ðe</i> <stœðe>		<i>*stø:ðe</i> <stœðe>	

Sample annotated data

(9) Examples from the etymologically annotated dataframe (abbreviated)

id	dipl	expanded	lemma	seq_no	v1	v2	etym1	etym2	v1_high	v2_high	VH
11857	stendr	stendr	standa	1	e	NA	ε	NA	F	NA	NA
30331	støðe	støðe	standa	1	ø	e	ø:	e	F	F	T
48957	stœpe	stœpe	standa	1	œ	e	ø:	e	F	F	T
65048	hofðíngíanom	hofðíngíanom	hofðingi	1	o	í	ɔ	i	F	T	F
65048	hofðíngíanom	hofðíngíanom	hofðingi	2	í	a	i	a	T	F	F
65048	hofðíngíanom	hofðíngíanom	hofðingi	3	a	o	a	o	F	F	T
34024	kkunar	kirkiunnar	kirkia	1	u	a	u	a	T	F	F
18773	giæva	giæva	giof	1	æ	a	æ	a	F	F	T
43804	giava	giava	giof	1	a	a	a	a	F	F	T

Outline

- 1 Background
 - Old Norwegian vowel harmony
 - Old Norwegian vowel harmony decay
- 2 Old Norwegian corpus and grapho-phonological methods
- 3 **Sanity check**
 - Lexical frequencies
 - Syllable and vowel frequencies
- 4 Tracking vowel harmony decay
 - Visualising decay
- 5 Conclusions

Outline

- 1 Background
 - Old Norwegian vowel harmony
 - Old Norwegian vowel harmony decay
- 2 Old Norwegian corpus and grapho-phonological methods
- 3 **Sanity check**
 - **Lexical frequencies**
 - Syllable and vowel frequencies
- 4 Tracking vowel harmony decay
 - Visualising decay
- 5 Conclusions

Lexical frequencies and Zipf's Law

Lexical frequencies and Zipf's Law

Annotation-code like (7) for the 600 most frequent lemmas in the corpus

Lexical frequencies and Zipf's Law

Annotation-code like (7) for the 600 most frequent lemmas in the corpus

- ▶ though only 7.2% of distinct lemmas, captures over 80% of word-tokens (185,534/222,082)

Lexical frequencies and Zipf's Law

Annotation-code like (7) for the 600 most frequent lemmas in the corpus

- ▶ though only 7.2% of distinct lemmas, captures over 80% of word-tokens (185,534/222,082)
 - ▶ Old Norwegian texts, like any natural language corpus, displays a zipfian distribution.

Lexical frequencies and Zipf's Law

Annotation-code like (7) for the 600 most frequent lemmas in the corpus

- ▶ though only 7.2% of distinct lemmas, captures over 80% of word-tokens (185,534/222,082)
 - ▶ Old Norwegian texts, like any natural language corpus, displays a zipfian distribution.

Zipf's law:

- ▶ the frequency of any word (lemma) is inversely proportional to its frequency rank (Zipf 1935)

Lexical frequencies and Zipf's Law

Annotation-code like (7) for the 600 most frequent lemmas in the corpus

- ▶ though only 7.2% of distinct lemmas, captures over 80% of word-tokens (185,534/222,082)
 - ▶ Old Norwegian texts, like any natural language corpus, displays a zipfian distribution.

Zipf's law:

- ▶ the frequency of any word (lemma) is inversely proportional to its frequency rank (Zipf 1935)
 - ▶ lemma frequencies cluster around zero

Lexical frequencies and Zipf's Law

Annotation-code like (7) for the 600 most frequent lemmas in the corpus

- ▶ though only 7.2% of distinct lemmas, captures over 80% of word-tokens (185,534/222,082)
 - ▶ Old Norwegian texts, like any natural language corpus, displays a zipfian distribution.

Zipf's law:

- ▶ the frequency of any word (lemma) is inversely proportional to its frequency rank (Zipf 1935)
 - ▶ lemma frequencies cluster around zero
 - ▶ e.g. 3,636 (44%) of lemmas in this corpus are *hapax legomena* – occurring only once in the entire corpus.

Lexical frequencies and Zipf's Law

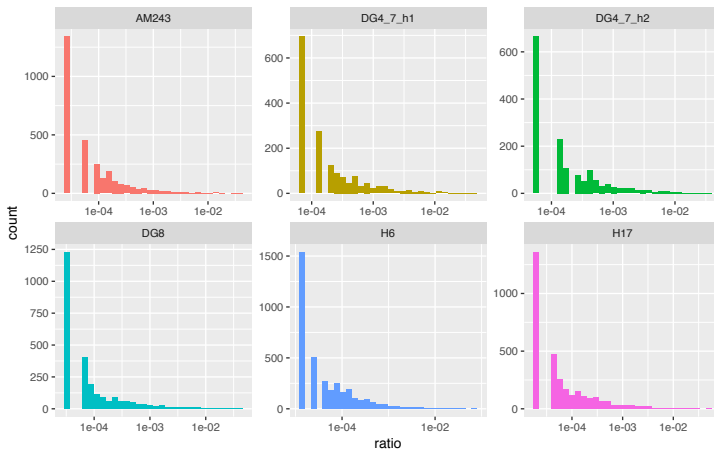


Figure 6: Lemma frequency ratios by manuscript

Lexical frequencies and Zipf's Law

Zipfian distributions are best observed by plotting the log frequency rank order by log lemma frequency, as in Fig. 7.

- ▶ spreading out the skewed distribution, straightening out the line.

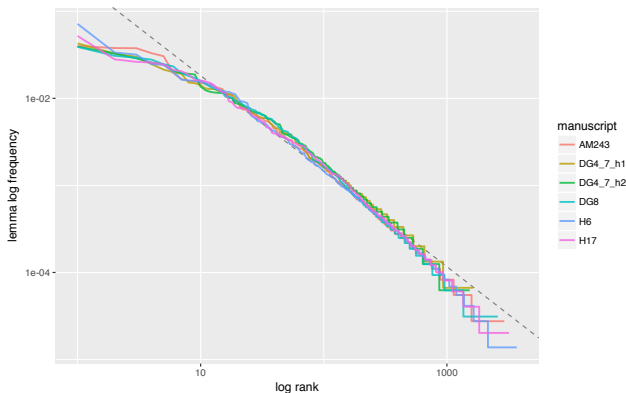


Figure 7: A log-log plot of lemma frequency rank by frequency across manuscripts

Outline

- 1 Background
 - Old Norwegian vowel harmony
 - Old Norwegian vowel harmony decay
- 2 Old Norwegian corpus and grapho-phonological methods
- 3 **Sanity check**
 - Lexical frequencies
 - **Syllable and vowel frequencies**
- 4 Tracking vowel harmony decay
 - Visualising decay
- 5 Conclusions

Syllable and vowel frequencies

The manuscripts display good agreement on vowel height class frequencies

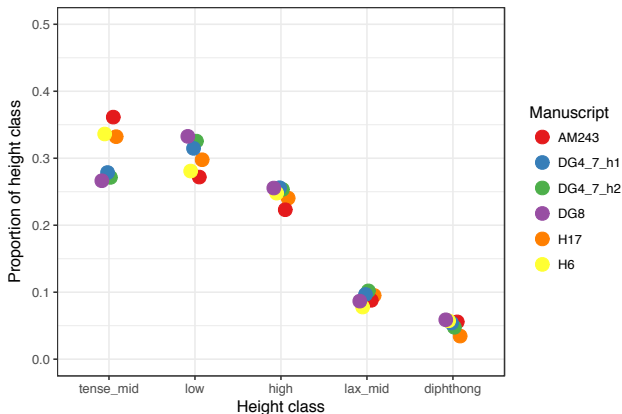


Figure 8: Height class proportions by manuscript

Syllable and vowel frequencies

The manuscripts display good agreement on average word length frequencies

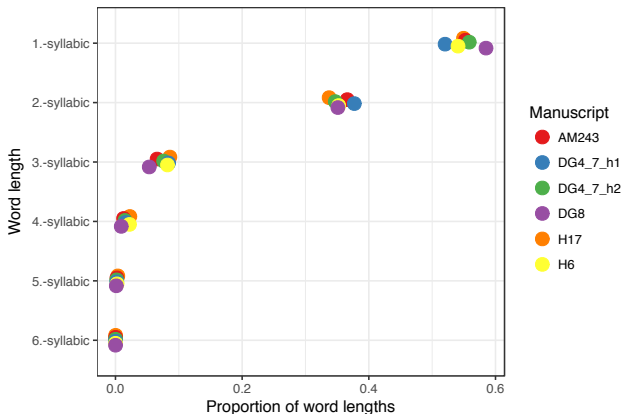


Figure 9: Word length frequencies across Old Norwegian manuscripts

Sanity check conclusions

- ▶ lexical frequencies suggest natural language vocabulary
 - ▶ lexical, syllable, and vocalic frequencies are uniform across the corpus
-  valid data sources

Outline

- 1 Background
 - Old Norwegian vowel harmony
 - Old Norwegian vowel harmony decay
- 2 Old Norwegian corpus and grapho-phonological methods
- 3 Sanity check
 - Lexical frequencies
 - Syllable and vowel frequencies
- 4 Tracking vowel harmony decay
 - Visualising decay
- 5 Conclusions

Comparing harmony frequencies across manuscripts

Goal: visualise and compare harmony frequencies across corpora

- ▶ tracking vowel harmony decay in Old Norwegian

PhonMatrix visualisations

PhonMatrix visualisations:

- ▶ developed by Mayer, Rohrdantz, et al. (2010) and Mayer & Rohrdantz (2013)
 - ▶ accessible at <http://phonmatrix.herokuapp.com/>

PhonMatrix visualisations

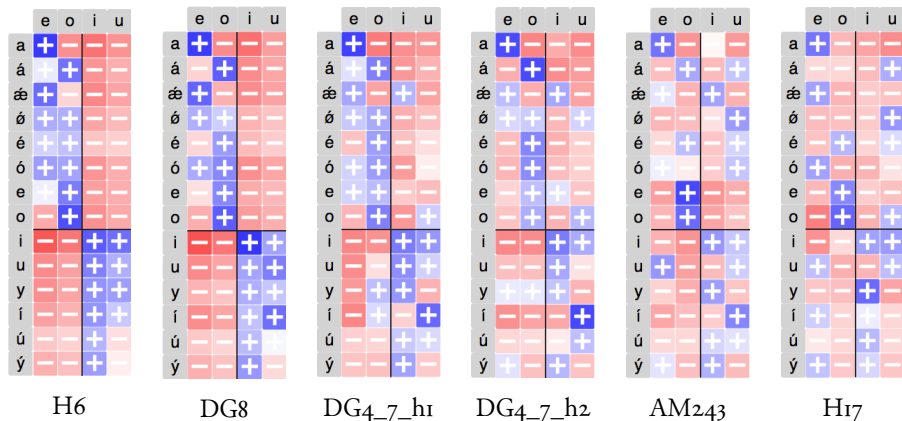


Figure 10: *PhonMatrix* visualisations of 13th-century Old Norwegian harmony decay

PhonMatrix visualisations

PhonMatrix takes as an input a V_1 – V_2 vowel matrix

- ▶ each vowel pair is assigned some association measure based on their frequency of occurrence
 - ▶ e.g. using the phi coefficient

(10) [a...e] contingency table

	[e]	not-e	Total
[a]	v	x	a
not-a	y	z	b
Total	c	d	

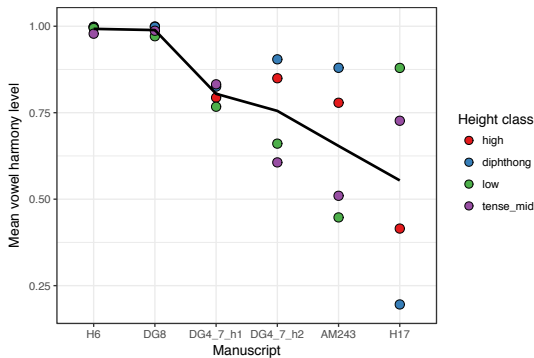
For the vowels in (10), the formula for the phi coefficient would be: $\phi = \frac{v \cdot z - x \cdot y}{\sqrt{a \cdot b \cdot c \cdot d}}$

PhonMatrix visualisations

The phi coefficient ranges from -1 to 1

- ▶ *PhonMatrix* visualisation maps the phi values to a bipolar colour scale (from red to blue)
- ▶ the darkness of the colour provides a visual indicator of the strength of each V_1 - V_2 association
 - ▶ positive associations are blue
 - ▶ negative associations are red

Distinct stages of vowel harmony decay



	Manuscript		Provenance	Date
1	H6	Holm perg 6 fol	Eastern Norway	c 1275
2	DG8	De la Gardie 8 fol, fol. 70v-110v	Trøndelag	c 1225-50
3	DG4_7_h1	De la Gardie 4-7, fol. 17v-29v	Bergen	c 1270
4	DG4_7_h2	De la Gardie 4-7, fol. 30r-43v	Bergen	c 1270
5	AM243	AM 243 b α fol	Bergen	c 1275
6	H17	Holm perg 17 4to	Uncertain	c 1300

Figure 11: Mean harmony levels by manuscript height class in pairwise sequences

Coherent directions of change

(II) Post-harmony inflectional systems

hu:s-i house-DAT.SG.

hu:s-um house-DAT.PL.

ljós-i light-DAT.SG.

ljós-um light-DAT.PL.

hús-e house-DAT.SG.

hús-om house-DAT.PL.

ljós-e light-DAT.SG.

ljós-om light-DAT.PL.

(a) AM243 – [-i] / [-um]

(b) Hr7 – [-e] / [-om]

hús-i house-DAT.SG.

hús-um house-DAT.PL.

ljós-i light-DAT.SG.

ljós-um light-DAT.PL.

hus-e house-DAT.SG.

hus-om house-DAT.PL.

dal-e valley-DAT.SG.

dal-om valley-DAT.PL.

(c) Mod.Icelandic – [-i] / [-um]

(d) Mod.Norw. (eastern) – [-e] / [-om]

Outline

- 1 Background
 - Old Norwegian vowel harmony
 - Old Norwegian vowel harmony decay
- 2 Old Norwegian corpus and grapho-phonological methods
- 3 Sanity check
 - Lexical frequencies
 - Syllable and vowel frequencies
- 4 Tracking vowel harmony decay
 - Visualising decay
- 5 Conclusions

Summary and conclusions

Vowel harmony decay: rarely attested and poorly understood sound change

- ▶ Old Norwegian provides us with rare and typologically significant insights
 - ▶ but comes with significant philological and historical phonological challenges

Summary and conclusions

Novel corpus linguistic methods:

- ▶ automated data collection using MENOTA transcriptions
- ▶ automated clean-up and linguistic annotation
- ▶ allows for easy stats/analysis

Summary and conclusions

Generalisations:

- ➊ Documented harmony decay
 - ▶ clear correlation between harmony levels and dispersion of height class behaviours
- ➋ Distinct stages of decay
 - ▶ *pre-decay*: H6 & DG8 – (high harmony ~ low variance)
 - ▶ *transitional*: DG4_7_h1/h2 – (medium harmony ~ medium variance)
 - ▶ *post-decay*: AM243 & H17 – (low harmony ~ high variance)
- ➌ Coherent directions of change
 - ▶ towards Icelandic-like and (eastern) Norwegian-like suffixes

Aims and directions for future research

Dialectology:

- ▶ Old Norwegian harmony decay provides rich variation
 - ▶ can be used to better map geographic and chronological variation

Harmony diagnostics:

- ▶ aids for identifying modern systems in the course of decay

Orthogonal sound processes/sound changes:

- ▶ These corpus methods allow for broader linguistic and philological study
 - ▶ e.g. sound mergers/splits, umlaut processes, scribal hand identification, etc.

Thanks for listening!

Cited works I

- Altan, Asli. 2007. The acquisition of vowel harmony in Turkish. Paper presented at the 4th Old World Conference in Phonology, University of the Aegean, 18–21 January.
- Bobaljik, Jonathan David. 2018. Disharmony and decay: Itelmen vowel harmony in the 20th century. In Roberto Petrosino, Pietro Cerrone & Harry van der Hulst (eds.), *From sounds to structures: Beyond the Veil of Maya*, 161–92. Berlin: Mouton de Gruyter.
- Csató, Éva Á. & Lars Johanson. 1998. Turkish. In Lars Johanson & Éva Á. Csató (eds.), *The Turkic languages*, chap. 11, 203–35. London: Routledge.
- Flom, George T. 1934. Vowel-harmony in Old Norse and Northern Aberdeenshire Scotch. *The Journal of English and Germanic Philology* 33(2). 178–93.
- Gallagher, Gillian. 2010. *The perceptual basis of long-distance laryngeal restrictions*. Cambridge, Massachusetts: MIT PhD thesis.

Cited works II

- Hagland, Jan Ragnar. 1978. A note on Old Norwegian vowel harmony. *Nordic Journal of Linguistics* 1(2). 141–147.
- Harrison, K. David, Mark Dras & Berk Kapicioglu. 2006. Agent-based modeling of the evolution of vowel harmony. In M. Hirotani (ed.), *Proceedings of the Northeast Linguistic Society*, vol. 32, 217–36.
- Hødnebø, Finn. 1977. Trykk–vokalharmoni–vokalbalanse. In Einar G. Pétursson & Jónas Kristjánsson (eds.), *Sjötíu ritgerðir helgaðar Jakobi Benediktssyni. 20 júlí 1977*, 375–83. Reykjavík: Stofnun Árna Magnússonar.
- Holthausen, Ferdinand. 1948. *Vergleichendes und etymologisches Wörterbuch des Altwestnordischen: Altnorwegisch–isländischen*. Göttingen: Vandenhoeck & Ruprecht.
- Kavitskaya, Darya. 2013. Segmental inventory and the evolution of harmony in Crimean Tatar. *Turkic languages* 17. 86–114.

Cited works III

- Leiwo, Matti, Pirjo Kulju & Katsura Aoyama. 2002. The acquisition of Finnish vowel harmony. In Anu Airola, Antti Arppe, Orvokki Heinämäki, Matti Miestamo, Kaius Sinnemäki, Urho Määttä, Jussi Niemi, Kari Pitkänen & Mickael Suominen (eds.), *A man of measure: Festschrift in honour of Fred Karlsson*, 149–61. Special supplement to the SKY Journal of Linguistics 19. The Linguistic Association of Finland.
- MacWhinney, Brian. 1978. The acquisition of morphophonology. *Monographs of the Society for Research in Child Development* 43(1/2). 1–123.
- Mailhot, Frédéric. 2010. *Modelling the acquisition and evolution of vowel harmony*. Ottawa, Ontario: Carleton University PhD thesis.
- Mayer, Thomas & Christian Rohrdantz. 2013. PhonMatrix: Visualizing co-occurrence constraints of sounds. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: Proceedings of the conference system demonstrations*, 73–78.
<http://www.aclweb.org/anthology/P13-4013>. Sofia, Bulgaria: Association for Computational Linguistics.

Cited works IV

- Mayer, Thomas, Christian Rohrdantz, Miriam Butt, Frans Plank & Daniel A. Keim. 2010. Visualizing vowel harmony. *Journal of Linguistic Issues in Language Technology* 4(2). <https://pdfs.semanticscholar.org/435e/32ab8792c67029977667b8e2afbf415a235a.pdf>, 1–33.
- McCollum, Adam G. 2015. Labial harmonic shift in Kazakh: Mapping the pathways and motivations for change. In *Proceedings of the 41st Annual Meeting of the Berkeley Linguistics Society*, 329–52.
- Ọla Orie, Ọlanikẹ. 2001. An alignment-based account of vowel harmony in Ifẹ Yoruba. *Journal of African Languages and Linguistics* 22(2). 117–43.
- Ọla Orie, Ọlanikẹ. 2003. Two harmony theories and high vowel patterns in Ebira and Yoruba. *The Linguistic Review* 20. 1–35.
- Ringen, Catherine O. 1975. *Vowel harmony: Theoretical implications*. Indiana University PhD thesis.

Cited works V

- Sandstedt, Jade J. 2017. Transparency and blocking in Old Norwegian height harmony. *Transactions of the Philological Society* 115(3). 395–417.
https://hcommons.org/app/uploads/sites/1000836/2019/03/thesis_final.pdf.
- Sandstedt, Jade J. 2018. *Feature specifications and contrast in vowel harmony: The orthography and phonology of Old Norwegian height harmony*. Online: <https://jsandstedt.hcommons.org/publications-presentations/>. University of Edinburgh PhD thesis.
- Seip, Didrik Arup. 1955. *Norsk språkhistorie til omkring 1370*. 2nd edn. Oslo: Aschehoug.
- Sjoberg, Andrée F. 1963. *Uzbek structural grammar*. Bloomington, Indiana: Indiana University Press.
- Suomi, Kari. 1983. Palatal harmony: A perceptually motivated phenomenon? *Nordic Journal of Linguistics* 6(1). 1–35.
- Walker, Rachel. 2005. Weak triggers in vowel harmony. *Natural Language & Linguistic Theory* 23(4). 917–89.

Cited works VI

Zipf, George Kingsley. 1935. *The psycho-biology of language*. Boston: Houghton Mifflin.