

Patternization – How to measure syntactic diversity?

Alexander Pfaff
alexander.pfaff@ling.uni-stuttgart.de
University of Stuttgart

May 4, 2024

Point of departure

*Constraints on Syntactic Variation: Noun Phrases in Early Germanic Languages**

(= NPEGL)

<https://www.hf.uio.no/ilos/english/research/projects/noun-phrases-in-early-germanic/>

- Research questions ↔ nomen est omen
- corpora, quantitative approaches
- two major outputs:
 - (i) database** specifically dedicated to noun phrases (NPEGL)
<https://spraakbanken.gu.se/en/resources/npegl>
 - (ii) edited volume
- highly specialized database allows detailed annotation
- how to best comply with the project title?

* @University of Oslo

** In collaboration with *Språkbanken, Göteborgs Universitet*; special thanks to Gerlof Bouma

Noun phrases in early Germanic languages

Edited by

Kristin Bech

Alexander Pfaff

Open Germanic Linguistics 8



Contents

Preface	iii
Acknowledgements	xi
1 The NPEGL noun phrase database: Design and construction Alexander Pfaff & Gerlof Bouma	1
2 How to measure syntactic diversity: Pattertnization, methods, algorithms Alexander Pfaff	33
3 Noun phrase modifiers in early Germanic: A comparative corpus study of Old English, Old High German, Old Icelandic, and Old Saxon Kristin Bech, Hannah Booth, Kersti Börjars, Tine Breban, Svetlana Petrova & George Walkden	71
4 The Old English quantifiers <i>fele</i> 'many' and <i>manig</i> 'many', and <i>Ælfric</i> as a linguistic innovator Kristin Bech	111
5 A new perspective on parallel inflection with reference to Old High German and Alemannic Alexandra Rehn	143
6 On the distribution of the strong and weak adjectival inflection in Old High German: A corpus investigation Svetlana Petrova	181
7 Beyond given versus new: The proprial article in Old Icelandic Hannah Booth	219
8 Modifying variation: Adjective position in Old Norwegian Juliane Tiemann	269

Entry in the NPEGL database

LANGUAGE	Old Icelandic		
DB ITEM ID	Olce.116.684		
CONTEXT	Ekki hafði hann mikka ást af fóður sínum. Engi var hann verklundarmaður. Vali hét maður er þar óx upp heima hjá Ófeigi. Hann var vænn maður og vinsæll . Oddur óx upp heima með fóður sínum þar til er hann var tólf vetra gamall.		
CORPUS UNIT ID	1350.BANDAMENNM.NAR-SAG.21		
GENDER	Mas		
NUMBER	Sg		
CASE	Nom		
GRAMMATICAL FUNCTION	Pred.Cop		
REFERENTIAL STATUS			
SEGMENTATION	[vænn] _{vænn} [maður] _{maður} [og] _{og} [vinsæll] _{vinsæll}		
ANNOTATION	vænn	Md.Aj.Lx.Pro	Eval, Str, Pos
	maður	N.C	Anim.HInd
	og	&Aj	
	vinsæll	Md.Aj.Lx.Pro	Eval, Str, Pos
DEGREE OF INTEREST	Green		
VERIFIED	Completely		
COMMENTS			
CONTAINS			
IS CONTAINED BY			
EXTERNAL LINKS			
ANNOTATOR	a.p.pfaff@ilos.uio.no		
ANNOTATION TIME	July 13th 2019, 19:44:10		

The annotation interface

Language Old Icelandic

DB item id Olce_042_086

Context Í flestum löndum setja menn á bækur annað tveggja þann frótleik er þar innan lands hefir gjörst – eða þann annan er minnisamlegastur þykir þó að annars staðar hafi heldur gjörst – | eða lög sín setja menn á bækur hver þjóð á sína tungu. | En af því að **tungumar** eru ólíkar hver annarri, **þær þegar er úr einni og hinní sömu tungu hafa gengist eða greinst**, þá þarf ólíka stafi í að hafa en eigi hina sömu alla í öllum, sem eigi rita grikkir latinustöfum gírskuna og eigi latinmenn gírskum stöfum latínu, né enn heldur hebreskir menn hebreskuna hvorki grískum stöfum né latínu. | heldur ritar sínum stöfum hver þjóð sína tungu.

Corpus unit id 1150.FIRSTGRAMMAR.SCI-LIN_3

 **Gender** Fem

 **Number** Pl

 **Case** Nom

 **Grammatical function** Arg.ofV.Sb

 **Referential status** Uniq

 **Segmentation** [tungumar]_{tungu} (eru ólíkar hver annarri.) [þær]_{só} [þegar er úr einni og hinní sömu tungu hafa gengist eða greinst]_s

Create Markable

Annotation  tungumar N.C Abst.Oth Def, Sf

 þær Dem Def

 þegar er úr einni og hinní sömu tungu hafa gengist eða greinst RC

 **Degree of interest** Green

The annotation interface: modifiers

✕ ✎ Segmentation [enskir]enskir [menn]maður

Create Markable

Annotation ✕ ✎ TEXT enskir

CAT Md.Aj.Lx.Pro

FEATURES [Core lexical] [Core Phrasal/Clausal] [Subdependent] INDICES [Coordination] TAGS ✕ ✎ menn N.C Anim.Hum

✕ ✎ Degree of interest Green

✕ ✎ Verified No

✕ ✎ Comments

(iii) Path Notation

(ii) Abbr.

The diagram illustrates the annotation interface for the word "enskir". The interface shows the word segmented into "enskir" and "maður" with the category "Md.Aj.Lx.Pro" selected. A dropdown menu for "Md.Aj.Lx.Pro" is open, showing a hierarchy of categories: Noun, Hinn, Pers, Dem, Poss, Quant, Mod, Posit, Num/WkQ, Adj, Func, Lex, PastP, PresP, Deriv, and Proto. Dashed lines connect the category labels to their corresponding items in the dropdown menu. The label "(iii) Path Notation" points to the "enskir" text, and the label "(ii) Abbr." points to the "Md.Aj.Lx.Pro" category.

The Modifier System

cat ⁰	cat ¹	cat ²	cat ³
Mod	. Posit		
	. Nu/WQ	. Num	
		. WQ	
	. Adj	. Func	. <i>Deter</i>
			. <i>Defect</i>
			. <i>Ord</i>
		. Lex	. <i>Proto</i>
			. <i>Deriv</i>
			. <i>PastP</i>
			. <i>PresP</i>

Table: (Adjectival) Modifier (Sub-)Categories

Nominal Properties

NounSem	:	Anim	.	HInd	Human-Individual
			.	HColl	Human-Collective
			.	Oth	Other (animate)
	:	Tang	.	Obj	Object
			.	Subs	Substance
	:	Abstr	.	Dyn	Dynamic entities
			.	Oth	Other (abstract)

Table: *Nominal Semantic Properties*

“Sf” = *suffixes article* (only relevant for the North Germanic languages)

“Def” = *definite* (annotated with N.P and N.C + “Sf”)

“Rel” = *relational noun*

The Pilot study (Bech at al.)

Pre- vs. postnominal modifier elements in Old Germanic languages

	Got	OE	OHG	OI	OS
Adj N	446 (53.3%)	42,375 (95.9%)	3,097 (81.7%)	3,529 (79.6%)	1,642 (84.2%)
N Adj	391 (46.7%)	1,821 (4.1%)	694 (18.3%)	904 (20.4%)	307 (15.8%)
CARD N	277 (82.9%)	8,075 (96.7%)	662 (90.9%)	616 (93.3%)	108 (79.4%)
N CARD	57 (17.1%)	278 (3.3%)	66 (9.1%)	44 (6.9%)	28 (20.6%)
POSS N	163 (12.1%)	29,647 (99.7%)	3,528 (82.0%)	1,339 (30.5%)	1,403 (93.7%)
N POSS	1,185 (87.9%)	78 (0.3%)	774 (18.0%)	3,057 (69.5%)	94 (6.3%)
PART N	13 (24.5%)	2,190 (92.1%)	176 (67.7%)	77 (87.5%)	64 (88.9%)
N PART	40 (75.5%)	189 (7.9%)	84 (32.3%)	11 (12.5%)	8 (11.1%)
QUANT N	292 (84.4%)	18,179 (97.6%)	1,350 (86.1%)	1,742 (84.8%)	261 (74.4%)
N QUANT	54 (15.6%)	442 (2.4%)	218 (13.9%)	312 (15.2%)	90 (25.6%)
MOD N	1,191 (40.8%)	100,466 (97.3%)	8,813 (82.8%)	7,303 (62.8%)	1,774 (80.6%)
N MOD	1,727 (59.2%)	2,808 (2.7%)	1,836 (17.2%)	4,328 (37.2%)	428 (19.4%)

The Pilot study (Bech at al.)

Pre- vs. postnominal modifier elements in Old Germanic languages

	Got	OE	OHG	OI	OS
ADJ N	446 (53.3%)	42,375 (95.9%)	3,097 (81.7%)	3,529 (79.6%)	1,642 (84.2%)
N ADJ	391 (46.7%)	1,821 (4.1%)	694 (18.3%)	904 (20.4%)	307 (15.8%)
CARD N	277 (82.9%)	8,075 (96.7%)	662 (90.9%)	616 (93.3%)	108 (79.4%)
N CARD	57 (17.1%)	278 (3.3%)	66 (9.1%)	44 (6.9%)	28 (20.6%)
POSS N	163 (12.1%)	29,647 (99.7%)	3,528 (82.0%)	1,339 (30.5%)	1,403 (93.7%)
N POSS	1,185 (87.9%)	78 (0.3%)	774 (18.0%)	3,057 (69.5%)	94 (6.3%)
PART N	13 (24.5%)	2,190 (92.1%)	176 (67.7%)	77 (87.5%)	64 (88.9%)
N PART	40 (75.5%)	189 (7.9%)	84 (32.3%)	11 (12.5%)	8 (11.1%)
QUANT N	292 (84.4%)	18,179 (97.6%)	1,350 (86.1%)	1,742 (84.8%)	261 (74.4%)
N QUANT	54 (15.6%)	442 (2.4%)	218 (13.9%)	312 (15.2%)	90 (25.6%)
MOD N	1,191 (40.8%)	100,466 (97.3%)	8,813 (82.8%)	7,303 (62.8%)	1,774 (80.6%)
N MOD	1,727 (59.2%)	2,808 (2.7%)	1,836 (17.2%)	4,328 (37.2%)	428 (19.4%)

⇒ ist there a more creative way to quantify syntactic diversity?

⇒ **Patternization!**

Patterns

- (1) The four patterns (Pfaff, 2014, 2015, 2016, 2017, 2019)
- a. **A-WK N -DEF** (I)
 gul-**i** bíll -**inn**
 yellow-WK car -DEF
- b. **ART A-WK N** (II)
hinn fullkornn-i glæpur
 ART perfect-WK crime
- c. **N -DEF A-WK** (III)
 heimspekingur -**inn** mikl-i
 philosopher -DEF great-WK
- d. **A-STR N -DEF** (IV)
 full-**ur** strákur -**inn**
 drunk-STR boy -DEF

Patterns

(1) The four patterns (Pfaff, 2014, 2015, 2016, 2017, 2019)

a. **A-WK N -DEF** (I)

gul-i bíll -inn
yellow-WK car -DEF

b. **ART A-WK N** (II)

hinn fullkomin-i glæpur
ART perfect-WK crime

c. **N -DEF A-WK** (III)

heimspekingur -inn mikl-i
philosopher -DEF great-WK

d. **A-STR N -DEF** (IV)

full-ur strákur -inn
drunk-STR boy -DEF

(i) ADJ N

(ix) Q ADJ N

(xv) POSS N RC

(ii) N ADJ

(x) Q DET ADJ N

(xvi) N DEM POSS ADJ

(iii) POSS ADJ N

(x- α) Q N DET A

.....

(iv) ADJ N POSS

(xi) ADJ POSS N

(???-a)

(v) N POSS ADJ

(xii) Q N

Patterns: Linear sequences of Cat labels

	<i>þessir</i>	<i>tveir</i>	<i>stóru</i>	<i>hestar</i>	
cat⁰	Dem	Md	Md	N	→ <i>patt⁰</i>
cat¹	Dem	Md.Card	Md.Aj	N.C	→ <i>patt¹</i>
cat²	Dem	Md.Card.Nu	Md.Aj.Lx	N.C	→ <i>patt²</i>
cat³	Dem	Md.Card.Nu	Md.Aj.Lx.Pro	N.C	→ <i>patt³</i>

Table: Pattern construal @ 4 levels of the NP “these two big horses”

patt⁰ : (Dem, Md, Md, N)

patterns are

patt¹ : (Dem, Md.Card, Md.Aj, N.C)

encoded

patt² : (Dem, Md.Card.Nu, Md.Aj.Lx, N.C)

as tuples

patt³ : (Dem, Md.Card.Nu, Md.Aj.Lx.Pro, N.C)

mixed : (Dem, Md, Md.Aj.Lx, N.C)

→ 2nd position underspecified

NPEGL: Basic Numbers

(ndb)		OldIcel	OldEngl	OldHighGer	OldSwed	OldSax
NPs		7981	3260	604	687	6696
CATs	cat ⁰	19	16	16	17	16
	cat ¹	25	22	20	21	20
	cat ²	28	27	23	24	23
	cat ³	34	30	28	31	28
PATTs	patt ⁰	384	151	92	75	245
	patt ¹	509	191	103	86	289
	patt ²	590	214	113	99	351
	patt ³	708	260	124	107	383

Table: *ndb*-subdatabases in NPEGL: NPs, categories, patterns

Pattern Diversity: Patterns per IXPs

	OldIcel	OldEngl	OldHighGer	OldSwed	OldSax
cat ⁰	4.8%	4.6%	15.2%	10.9%	3.7%
cat ¹	6.4%	5.9%	17.1%	12.5%	4.3%
cat ²	7.4%	6.6%	18.7%	14.4%	5.2%
cat ³	8.9%	8.0%	20.5%	15.6%	5.7%

Table: Pattern Diversity: patterns per NPs

$$(1) \quad \mu\text{-}p = \frac{1}{n} \sum_{i=1}^n p_i \quad \Rightarrow \quad \text{PATTDIV} = \frac{\mu\text{-}p}{\text{scd}}$$

(with p_i = number of patterns in sub-database rnd_DB_i)

	OldIcel	OldEngl	OldHighGer	OldSwed	OldSax
cat ⁰	13.1%	8.6%	15.2%	10.9%	9.5%
cat ¹	16.5%	10.6%	17.1%	12.5%	10.2%
cat ²	18.5%	11.7%	18.7%	14.4%	12.6%
cat ³	21.8%	13.8%	20.5%	15.6%	13.4%

Table: PATTDIV with $\text{scd} = 1000$; $n = 500$

Cool data from Old Icelandic

- (2) a. **sína fullkomna vináttu**
 POSS *perfect* *friendship*
- b. **fullkomna vináttu sína**
perfect *friendship* POSS
- c. **vináttu sinni fullkominni**
friendship POSS *perfect*
- d. **fullkominni sinni vináttu**
perfect POSS *friendship*
- e. **sinni vináttu fullkominni**
 POSS *friendship* *perfect*
 “his perfect/complete friendship”

⇒ 5 patterns involving the categories N, ADJ, POSS

⇒ Potentially, $3! = 6$ distinct permutations of 3 objects

Combinatorial Flexibility

⇒ Attestation as a binary parameter

	⇒	{N, POSS, ADJ}		
	i.	POSS ADJ N	[+ATT]	1 / True
	ii.	ADJ N POSS	[+ATT]	1 / True
(3)	iii.	N POSS ADJ	[+ATT]	1 / True
	iv.	ADJ POSS N	[+ATT]	1 / True
	v.	POSS N ADJ	[+ATT]	1 / True
	vi.	N ADJ POSS	[-ATT]	0 / False

⇒ **COMBFLEX**({N, ADJ, POSS}) = 5/6

Basic Combinatorics

Given a *sample space* S , with $|S| = n$, and $k \in \mathbb{N} \leq n$, then there are

- $n! = n \times (n-1) \times (n-2) \times \dots \times 2 \times 1$ (full) **permutations** (of n)
- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ **k -combinations** (of n) \Rightarrow sub-sets of size k
- $\binom{n}{k} \times k! = \frac{n!}{(n-k)!}$ **k -permutations** (of n) \Rightarrow sub-permutations of size k

Assume *sample space* $S = \{A, B, C, D, E, F\}$ with $n = |S| = 6$; let $k = 3$; then there are

$$\binom{6}{3} = \frac{6!}{3!(6-3)!} = \mathbf{20}$$

different (subsets of size 3 =) **3-combinations** (of 6):

{A, C, B}	{A, C, D}	{A, D, E}	{A, C, E}	{A, B, D}
{A, D, F}	{A, E, F}	{A, C, F}	{A, E, B}	{A, B, F}
{F, D, C}	{F, E, B}	{B, C, D}	{C, E, B}	{B, C, F}
{B, D, F}	{C, D, E}	{B, D, E}	{C, E, F}	{D, E, F}

NB: {A, C, B} == {C, A, B} == {B, A, C} == {B, C, A} etc.

Basic Combinatorics

$\binom{6}{3} = 20 \times 3! = 6 = 120$ ways to select **3-permutations** (of 6):

(A, C, B)	(A, B, C)	(B, A, C)	(B, C, A)	(C, A, B)	(C, B, A)
(A, C, D)	(A, D, C)	(C, A, D)	(C, D, A)	(D, A, C)	(D, C, A)
(A, D, E)	(A, E, D)	(D, A, E)	(D, E, A)	(E, A, D)	(E, D, A)
(A, C, E)	(A, E, C)	(C, A, E)	(C, E, A)	(E, A, C)	(E, C, A)
(A, B, D)	(A, D, B)	(B, A, D)	(B, D, A)	(D, A, B)	(D, B, A)
(A, D, F)	(A, F, D)	(D, A, F)	(D, F, A)	(F, A, D)	(F, D, A)
(A, E, F)	(A, F, E)	(E, A, F)	(E, F, A)	(F, A, E)	(F, E, A)
(A, C, F)	(A, F, C)	(C, A, F)	(C, F, A)	(F, A, C)	(F, C, A)
(A, E, B)	(A, B, E)	(B, A, E)	(B, E, A)	(E, A, B)	(E, B, A)
(A, B, F)	(A, F, B)	(B, A, F)	(B, F, A)	(F, A, B)	(F, B, A)
(F, D, C)	(C, F, D)	(D, C, F)	(D, F, C)	(F, C, D)	(C, D, F)
(F, E, B)	(B, F, E)	(E, B, F)	(E, F, B)	(F, B, E)	(B, E, F)
(B, C, D)	(B, D, C)	(C, B, D)	(C, D, B)	(D, B, C)	(D, C, B)
(C, E, B)	(B, E, C)	(C, B, E)	(B, C, E)	(E, B, C)	(E, C, B)
(B, C, F)	(B, F, C)	(C, B, F)	(C, F, B)	(F, B, C)	(F, C, B)
(B, D, F)	(B, F, D)	(D, B, F)	(D, F, B)	(F, B, D)	(F, D, B)
(C, D, E)	(C, E, D)	(D, C, E)	(D, E, C)	(E, C, D)	(E, D, C)
(B, D, E)	(B, E, D)	(D, B, E)	(D, E, B)	(E, B, D)	(E, D, B)
(C, E, F)	(C, F, E)	(E, C, F)	(E, F, C)	(F, C, E)	(F, E, C)
(D, E, F)	(D, F, E)	(E, D, F)	(E, F, D)	(F, D, E)	(F, E, D)

NB: $(A, C, B) \neq (C, A, B) \neq (B, C, A) \dots \neq \{A, B, C\}$

Relevant sample spaces & how to use them

$S_{cat} = \text{CAT} = \text{cat}^0 \cup \text{cat}^1 \cup \text{cat}^2 \cup \text{cat}^3$ (complete category set)

$S_{cat} = \text{cat}^1 / \text{cat}^2$ ($\text{cat}^1 / \text{cat}^2$ categories)

$S_{cat} = \{\text{Adv, Md.Aj, PP, Q, N, Dem, GenP, N.C, RC}\}$ (random selection)

$S_{cat} = \{\text{Md, Md.Card, Md.Aj, Md.Aj.Lx, Md.Aj.Fn, Md.Aj.Lx.Pro}\}$ (modifier selection)

$S_{cat} = \text{cat}^2 \cup \{\text{Md.Aj.Fn, Md.Aj.Lx.Pro}\}$ (mixed selection)

⇒ **CombFlex**(self, *cats, length=3, restrict="N.C", count=bool, sm_pattern=None, threshold=1, group_threshold=2, alignment=None):

**cats*: the sample space

length: length of sub-permutations over sample space

threshold: minimal number of occurrences to be considered +ATT

restrict: if the combination contains "N.C", the respective (sub-)permutations will be generated

func: function to diagnose M-patterns (see next slide)

count: `boolean` (= attestation: yes/no). or `integer` (= actual occurrences)

⇒ creates all sub-permutations of length *lang* in **cats*, if they contain "N.C",

⇒ then browses the database to check whether those patterns are attested

⇒ returns the permutation groups that satisfy the threshold conditions + count

S-search and M-match Patterns

⇒ S-Patterns

→ $F(\text{patt})$

■ precise_pattern:

A, B, C

⇒

| A, B, C |

■ rigid_pattern:

A, B, C

⇒

| ... A, B, C ... |

■ flexi_pattern:

A, B, C

⇒

| ... A, ..., B, ..., C ... |

■ Left_rigid_pattern:

A, B, C

⇒

| A, B, C ... |

■ Left_flexi_pattern:

A, B, C

⇒

| A, ..., B, ..., C ... |

■ Right_rigid_pattern:

A, B, C

⇒

| ... A, B, C |

■ Right_flexi_pattern:

A, B, C

⇒

| ... A, ..., B, ..., C |

M-Patterns ←

match/result ←

Some results in the Old Icelandic *ndb*

(4) a. {Md.Aj, App, N}:	1 / 6	d. {N, Md.Card.WQ, Md.Aj}:	4 / 6
i. (App, Md.Aj, N):	FALSE	i. (Md.Aj, Md.Card.WQ, N):	TRUE
ii. (App, N, Md.Aj)	FALSE	ii. (Md.Aj, N, Md.Card.WQ):	TRUE
iii. (Md.Aj, App, N):	FALSE	iii. (Md.Card.WQ, Md.Aj, N):	TRUE
iv. (Md.Aj, N, App):	TRUE	iv. (Md.Card.WQ, N, Md.Aj):	TRUE
v. (N, App, Md.Aj):	FALSE	v. (N, Md.Aj, Md.Card.WQ):	FALSE
vi. (N, Md.Aj, App):	FALSE	vi. (N, Md.Card.WQ, Md.Aj):	FALSE
b. {N, Dem, RC}:	2 / 6	e. {N, Md.Aj, Poss}:	5 / 6
i. (Dem, N, RC):	TRUE	i. (Md.Aj, N, Poss):	TRUE
ii. (N, Dem, RC):	TRUE	ii. (Poss, Md.Aj, N):	TRUE
iii. (RC, N, Dem):	FALSE	iii. (N, Poss, Md.Aj):	TRUE
iv. (RC, Dem, N):	FALSE	iv. (Poss, N, Md.Aj):	TRUE
v. (N, RC, Dem):	FALSE	v. (Md.Aj, Poss, N):	TRUE
vi. (Dem, RC, N):	FALSE	vi. (N, Md.Aj, Poss):	FALSE
c. {N, Dem, Md.Aj.Lx}:	3 / 6	f. {Q, N, Md.Aj}:	6 / 6
i. (Dem, Md.Aj, N):	TRUE	i. (Q, Md.Aj, N):	TRUE
ii. (Dem, N, Md.Aj):	TRUE	ii. (Q, N, Md.Aj):	TRUE
iii. (Md.Aj, N, Dem):	FALSE	iii. (Md.Aj, Q, N):	TRUE
iv. (N, Dem, Md.Aj):	TRUE	iv. (Md.Aj, N, Q):	TRUE
v. (Md.Aj, Dem, N):	FALSE	v. (N, Md.Aj, Q):	TRUE
vi. (N, Md.Aj, Dem):	FALSE	vi. (N, Q, Md.Aj):	TRUE

Some results may not be exciting, but the method as such is completely exhaustive

Mean combinatorial flexibility

	OldIcel	OldEngl	OldHighGer	OldSwed	OldSax
1/6	31	41	27	31	13
2/6	59	41	19	17	28
3/6	20	10	6	7	20
4/6	13	2	0	3	10
5/6	19	5	0	0	4
6/6	11	0	0	0	6

Table: Combinatorial flexibility in $S_{cat} = cat^2$ with $k = 3$

	OldIcel	OldEngl	OldHighGer	OldSwed	OldSax
μ -COMBFLEX	2.8/6	1.9/6	1.6/6	1.7/6	2.8/6

Table: Mean combinatorial flexibility in $S_{cat} = cat^2$ with $k = 3$

Modified mean combinatorial flexibility

C_{att} (= “attested combination”), if COMBFLEX 1/6 – 6/6 (= at least 1 x +ATT)

C_{pot} (= “potential combination”) = all valid combinations over the sample space¹

	OldIcel	OldEngl	OldHighGer	OldSwed	OldSax
categories	28	27	23	24	23
C_{pot}	351	325	231	253	231
C_{att}	153	99	52	58	81
$R_{pot}^{att} = \frac{C_{att}}{C_{pot}}$	0.436	0.305	0.225	0.229	0.351
$\mu\text{-COMBFLEX}_{pot}$	1.2/6	0.6/6	0.4/6	0.4/6	1.0/6

Table: Potential and attested combinations; modified combinatorial flexibility

¹For instance, Old Icelandic has 28 ATTESTED cat² categories, which includes the label “N”. Since, by stipulation, every combination must contain a noun, the procedure will generate $\binom{28-1}{2} = 351$ permutation groups, i.e. 3-combinations that contain “N”.

Combinatorial flexibility as indicator of markedness

	OldIcel	OldEngl	OldHighGer	OldSwed	OldSax
{Q, Poss, N}	5/6	1/6	0/6	3/6	3/6
(Q, Poss, N)	TRUE	TRUE	FALSE	TRUE	TRUE
(Q, N, Poss)	TRUE	FALSE	FALSE	TRUE	TRUE
(Poss, N, Q)	TRUE	FALSE	FALSE	TRUE	TRUE
(Poss, Q, N)	TRUE	FALSE	FALSE	FALSE	FALSE
(N, Poss, Q)	TRUE	FALSE	FALSE	FALSE	FALSE
(N, Q, Poss)	FALSE	FALSE	FALSE	FALSE	FALSE
{Md.Aj, Poss, N}	5/6	1/6	2/6	3/6	2/6
(Poss, Md.Aj, N)	TRUE	TRUE	TRUE	TRUE	TRUE
(Poss, N, Md.Aj)	TRUE	FALSE	TRUE	TRUE	TRUE
(N, Poss, Md.Aj)	TRUE	FALSE	FALSE	TRUE	FALSE
(Md.Aj, N, Poss)	TRUE	FALSE	FALSE	FALSE	FALSE
(Md.Aj, Poss, N)	TRUE	FALSE	FALSE	FALSE	FALSE
(N, Md.Aj, Poss)	FALSE	FALSE	FALSE	FALSE	FALSE

Table: COMBFLEX({Poss, N, Q}/{Poss, N, Md.Aj}) in comparison

CATs = {Poss, Q, Md.Aj.Lx, Md.Aj.Fn, Md.Nu/WQ.Nu, Md.Nu/WQ.WQ, GenP}

$\forall \text{ cat} \in \text{CATs:}$

$\forall \text{ np} \in \text{ndb:}$

- $\rightarrow \text{Right_flexi_pattern}(\text{np}, \text{"N.C"}, \text{cat}, \text{"Dem"})$

Poss	0
Q	0
Md.Aj.Lx	0
Md.Aj.Fn	0
Md.Nu/WQ.Nu	0
GenP	0

= [NP ... N.C ... cat ... Dem]

NB: [NP ... N.C Dem] \rightarrow 52

- $\rightarrow \text{flexi_pattern}(\text{np}, \text{"N.C"}, \text{cat}, \text{"Dem"})$

Poss	21
Q	10
Md.Aj.Lx	5
Md.Aj.Fn	1
Md.Nu/WQ.Nu	9
GenP	21

= [NP ... N.C ... cat ... Dem ...]

- $\rightarrow \text{flexi_pattern}(\text{np}, \text{"N.C"}, \text{cat}, \text{"Dem"}, \text{"RC"})$

Poss	20
Q	10
Md.Aj.Lx	5
Md.Aj.Fn	1
Md.Nu/WQ.Nu	9
GenP	19

= [NP ... N.C ... cat ... Dem ... **RC** ...]

\Rightarrow comparison of alignment D-pattern vs. general D-pattern to probe for correlations
(e.g. postnominally, Dem (sá) follows cat \in CATs only if accompanied by a RC)

Combinatorial Flexibility: { N.C, Md.Aj, **X** } — for $X = \{\text{Dem, Num, WQ, Poss, Q}\}$ (5) a. (**X**, Md.Aj, N.C)

i. (Dem, Md.Aj, N.C):	1	85	6
ii. (Md.Nu/WQ.Nu, Md.Aj, N.C):	1	31	6
iii. (Md.Nu/WQ.WQ, Md.Aj, N.C):	1	34	6
iv. (Poss, Md.Aj, N.C):	1	32	5
v. <u>(Q, Md.Aj, N.C):</u>	1	65	6
	5	247	5.8

b. (**X**, N.C, Md.Aj)

i. (Dem, N.C, Md.Aj):	1	05	4
ii. (Md.Nu/WQ.Nu, N.C, Md.Aj):	1	09	4
iii. (Md.Nu/WQ.WQ, N.C, Md.Aj):	1	12	5
iv. (Poss, N.C, Md.Aj):	1	03	2
v. <u>(Q, N.C, Md.Aj):</u>	1	14	5
	5	043	4.0

c. (Md.Aj, N.C, **X**)

i. (Md.Aj, N.C, Dem):	1	68	5
ii. (Md.Aj, N.C, Md.Nu/WQ.Nu):	1	04	2
iii. (Md.Aj, N.C, Md.Nu/WQ.WQ):	0	00	1
iv. (Md.Aj, N.C, Poss):	1	53	6
v. <u>(Md.Aj, N.C, Q):</u>	1	04	3
	4	129	3.4

d. (N.C, **X**, Md.Aj)

i. (N.C, Dem, Md.Aj):	1	02	2
ii. (N.C, Md.Nu/WQ.Nu, Md.Aj):	1	13	5
iii. (N.C, Md.Nu/WQ.WQ, Md.Aj):	0	00	1
iv. (N.C, Poss, Md.Aj):	1	17	4
v. <u>(N.C, Q, Md.Aj):</u>	1	05	4
	4	037	3.2

e. (Md.Aj, **X**, N.C)

i. (Md.Aj, Dem, N.C):	1	01	1
ii. (Md.Aj, Md.Nu/WQ.Nu, N.C):	1	07	3
iii. (Md.Aj, Md.Nu/WQ.WQ, N.C):	1	02	4
iv. (Md.Aj, Poss, N.C):	1	07	3
v. <u>(Md.Aj, Q, N.C):</u>	1	01	1
	5	018	2.4

f. (N.C, Md.Aj, **X**)

i. (N.C, Md.Aj, Dem):	1	04	3
ii. (N.C, Md.Aj, Md.Nu/WQ.Nu):	0	00	1
iii. (N.C, Md.Aj, Md.Nu/WQ.WQ):	0	00	1
iv. (N.C, Md.Aj, Poss):	0	00	1
v. <u>(N.C, Md.Aj, Q):</u>	1	02	2
	2	006	1.6

Co-Distribution

- (6) a. (\mathbf{x} , Md.Aj, N.C): 5
 b. (\mathbf{x} , N.C, Md.Aj): 5
 c. (Md.Aj, \mathbf{x} , N.C): 5
 d. (Md.Aj, N.C, \mathbf{x}): 4
 e. (N.C, \mathbf{x} , Md.Aj): 4
 f. (N.C, Md.Aj, \mathbf{x}): 1

$x \in \{WQ, \text{Dem, Num, Poss, } Q\}$

$x \in \{WQ, \text{Dem, Num, Poss, } Q\}$

$x \in \{WQ, \text{Dem, Num, Poss, } Q\}$

$x \in \{\text{Dem, Num, Poss, } Q\}$

$x \in \{\text{Dem, Num, Poss, } Q\}$

$x \in \{Q\}$

⇒ Dem, Poss and Num have an identical distribution in the context of N & A:

- all three occur in the permutations (6a-e),
- all three do not occur in the permutation (6f)

Aligning & stacking NPs

1	2	3	4	5
Dem	Adj	N		
Adj	N			
Dem	N	RC		
Q	Adj	N	Dem	RC
N	Adj	Dem	RC	
Dem	Adj	N		
Adj	N	GenP		
Num	Adj	N	GenP	
Dem	Num	N		
Dem	N	RC		
...

-5	-4	-3	-2	-1
		Dem	Adj	N
			Adj	N
		Dem	N	RC
Q	Adj	N	Dem	RC
	N	Adj	Dem	RC
		Dem	Adj	N
		Adj	N	GenP
	Num	Adj	N	GenP
		Dem	Num	N
		Dem	N	RC
...

The Proto NP — a probabilistic model

<p>N.C. : 2318 Dem : 900 Md.Aj.Lx : 870 Q : 679 H : 477 Poss : 406 Md.Aj.Fn : 378 Md.Nu/WQ.Nu : 291 GenP : 183 Md.Nu/WQ.WQ : 167 Md : 152 Mdmd : 139 Md.Pos : 58 Per : 21 Dgcm : 1 Md.Aj : 4 Md.Nu/WQ : 4 BP : 1 Adv : 1 Dgcm.Br : 2 Mdcm.N : 1 App : 1</p>	<p>N.C. : 3699 Poss : 1011 GenP : 679 Md.Aj.Lx : 545 Dgcm : 306 Md.Aj.Fn : 306 Md : 211 Md.Nu/WQ.Nu : 108 H : 100 Q : 81 Md.Nu/WQ.WQ : 46 Md.Aj : 2 RC : 4 Dgcm.Br : 1 App : 1 BP : 1 Mdmd : 13 Per : 2 PP : 1 Md.Pos : 1 Adv : 1</p>	<p>N.C. : 1009 RC : 582 Dem : 159 GenP : 140 Md.Aj.Lx : 124 Poss : 104 PP : 80 CC.FI : 68 Dgcm.Mk : 67 Q : 65 Md.Aj.Fn : 49 H : 32 Md : 30 App : 25 Disco : 23 Md.Nu/WQ.Nu : 2 Adv : 20 Mdcm.P : 1 CC.NF : 15 BP : 1 Mdmd : 1 Mdcm.N : 1 Md.Nu/WQ.WQ : 1 Assoc : 1</p>	<p>RC : 308 N.C. : 112 Dgcm.Mk : 71 GenP : 54 PP : 39 Md.Aj.Lx : 38 Dem : 54 App : 25 Assoc : 1 Dgcm.Br : 1 H : 4 Poss : 16 Q : 1 Adv : 1 Md.Aj.Fn : 1 Md : 1 Mdcm.P : 2 Mdcm.N : 1 CC.NF : 2 Mdmd : 1 CC.FI : 1 Md.Nu/WQ.Nu : 1 Mdcm.N : 4 Disco : 1</p>	<p>RC : 75 App : 11 Md.Aj.Fn : 1 Md : 1 Mdcm.P : 1 N.C. : 5 Mdmd : 2 Md.Nu/WQ.WQ : 1 Dgcm.Br : 1 GenP : 2 Poss : 2 Dgcm.Mk : 1 PP : 1 Md.Aj.Lx : 1 Dem : 1 Assoc : 1</p>	<p>RC : 8 Dgcm.Mk : 4 CC.FI : 1 App : 1 CC.NF : 1 Dem : 1 Disco : 1</p>	<p>RC : 1</p>	7143	7143	2663	772	134	17	1
1	17	134	772	2663	7143	7143							
<p>Md.Nu/WQ.Nu : 1</p>	<p>H : 2 N.C. : 2 Md.Nu/WQ.Nu : 4 Mdmd : 1 Md.Nu/WQ.WQ : 1 Dem : 2 Q : 1 Md.Aj.Lx : 2 Poss : 2</p>	<p>Per : 1 Md : 1 Poss : 1 Md.Aj.Fn : 6 Md.Nu/WQ.WQ : 7 Md.Aj.Lx : 6 Mdmd : 11 Md.Nu/WQ.Nu : 12 H : 15 Q : 15 N.C. : 21 Dem : 34</p>	<p>Dgcm : 1 GenP : 7 Md.Pos : 2 Adv : 4 Md : 8 Poss : 14 Per : 5 Md.Nu/WQ.WQ : 26 Md.Nu/WQ.Nu : 30 Md.Aj.Fn : 39 Mdmd : 57 Md.Aj.Lx : 67 R : 308 N.C. : 121 H : 133 Dem : 151</p>	<p>Dgcm : 1 GenP : 7 Md.Pos : 2 Adv : 4 Mdcm.N : 3 Poss : 14 Per : 5 Md.Nu/WQ.WQ : 26 Md.Nu/WQ.Nu : 30 Md.Aj.Fn : 39 Mdmd : 57 Md.Aj.Lx : 67 R : 308 N.C. : 121 H : 133 Dem : 151</p>	<p>Md.Pos : 13 RC : 5 Per : 6 PP : 2 Mdcm.N : 3 Md.Aj : 2 Md.Aj : 2 Dgcm.Br : 2 Mdmd : 82 Md.Nu/WQ.WQ : 71 Md : 68 Poss : 68 GenP : 49 Md.Nu/WQ.Nu : 114 Md.Aj.Fn : 201 Q : 210 Md.Aj.Lx : 231 H : 421 Dem : 88 N.C. : 522</p>	<p>Assoc : 4 Dgcm.Br : 1 Dgcm.Mk : 2 Adv : 4 Mdcm.P : 3 Md.Aj : 2 Md.Aj : 2 CC.NF : 1 Md.Nu/WQ : 5 BP : 1 Mdmd : 12 Per : 18 Mdcm.N : 5 App : 4 PP : 21 RC : 34 Md.Nu/WQ.WQ : 104 GenP : 213 Md.Nu/WQ.Nu : 223 Md : 229 Q : 417 Poss : 458 Md.Aj.Fn : 471 Dem : 761 Md.Aj.Lx : 1132 N.C. : 2954</p>	<p>Assoc : 18 Md.Pos : 3 Md.Nu/WQ.WQ : 13 Dgcm.Br : 2 H : 1 Mdcm.P : 1 CC.NF : 17 BP : 11 Mdmd : 5 Per : 1 Mdcm.N : 2 Adv : 23 Disco : 28 Md.Aj.Fn : 31 Md.Nu/WQ.Nu : 39 Dem : 52 App : 65 Q : 75 CC.FI : 78 Md : 98 PP : 116 Dgcm.Mk : 148 Md.Aj.Lx : 160 GenP : 789 RC : 941 Poss : 992 N.C. : 3422</p>						

The Proto NP – a probabilistic model; postnominal field

	Poss': 1090, 'GenP': 845, RC: 455, 'Dem': 413, 'Md.Aj.Lx': 139, 'Dgcm.Mk': 117, PP: 108, Q: 101, 'Md': 90, 'Md.Nu/WQ.Nu': 79, 'CC.F': 67, 'Md.Aj.Fn': 31, 'H': 28, 'disco': 26, 'App': 26, 'Adv': 20, 'Mdmd': 17, 'CC.NF': 14, 'Md.Nu/WQ.WQ': 11, 'IXP': 11, 'Mdcm.P': 10, 'Assoc': 10, 'Mdcm.N': 6, 'Md.Pos': 3, 'Dgcm.Br': 2, 'Per': 2	RC: 433, 'Dem': 54, 'Md.Aj.Lx': 53, Q: 28, 'App': 27, PP: 26, 'Dgcm.Mk': 25, 'H': 22, 'Md.Aj.Fn': 20, 'Md': 15, 'CC.F': 11, 'Assoc': 9, 'Mdmd': 9, 'Adv': 7, 'Mdcm.P': 7, 'Md.Nu/WQ.Nu': 7, 'GenP': 5, 'Md.Nu/WQ.WQ': 3, 'Mdcm.N': 3, 'CC.NF': 2, 'disco': 1	RC: 78, 'Md.Aj.Lx': 16, 'App': 12, 'Md': 9, 'Dgcm.Mk': 6, PP: 4, 'Dem': 3, 'Md.Aj.Fn': 3, 'Mdcm.P': 3, 'Assoc': 2, 'H': 2, 'CC.NF': 2, Q: 1, 'Dgcm.Br': 1, 'Mdmd': 1, 'Md.Nu/WQ.WQ': 1	RC: 12, 'App': 3, 'Md.Aj.Lx': 2, 'Dgcm.Mk': 2, 'Md.Aj.Fn': 1, 'Mdmd': 1, PP: 1, 'Assoc': 1	'disco': 1, RC: 1
N	3721	767	144	23	2
	2	23	144	767	3721
	'Md.Nu/WQ.Nu': 1, 'Poss': 1	'Poss': 4, 'Mdmd': 3, 'GenP': 2, 'H': 1, 'Md.Nu/WQ.Nu': 1, 'Md.Aj.Lx': 1, 'Dem': 11	'Mdmd': 4, RC: 4, 'Mdcm.N': 3, PP: 2, 'Assoc': 2, 'Md.Aj.Fn': 2, 'Mdcm.N': 2, 'Per': 1, 'Md.Aj.Lx': 10, 'H': 12, 'CC.NF': 3, Q: 11, 'Md.Nu/WQ.Nu': 16, 'GenP': 17, 'Poss': 21, 'Dem': 36	'Adv': 5, 'Mdcm.N': 5, 'Assoc': 4, 'App': 3, 'Md.Nu/WQ.WQ': 3, 'Mdcm.P': 3, 'Dgcm.Mk': 2, 'Dgcm.Br': 2, 'CC.NF': 1, 'Mdmd': 1, 'Md': 16, 'Md.Aj.Fn': 21, 'Md.Nu/WQ.Nu': 29, RC: 34, 'H': 38, 'Md.Aj.Lx': 39, 'GenP': 42, Q: 44, 'Poss': 72, 'Dem': 371	'Mdmd': 5, 'Md.Pos': 3, 'Mdcm.N': 2, 'Dgcm.Br': 2, 'Per': 1, 'IXP': 11, 'Md.Nu/WQ.WQ': 12, 'Assoc': 16, 'CC.NF': 17, 'Mdcm.P': 17, 'Adv': 22, 'disco': 28, 'Md.Aj.Fn': 31, 'Md.Nu/WQ.Nu': 39, 'Dem': 52, 'App': 65, Q: 75, 'CC.F': 78, 'Md': 98, PP: 116, 'Dgcm.Mk': 148, 'Md.Aj.Lx': 160, 'GenP': 789, RC: 941, 'Poss': 992

The Proto NP – a probabilistic model; prenominal field

'Dem': 980, 'Md.Aj.Lx': 870, Q': 679, H': 477, Poss: 406, 'Md.Aj.Fn': 378, 'Md.Nu/WQ.Nu': 291, 'GenP': 183, 'Md.Nu/WQ.WQ': 167, 'Md': 152, 'Mdmd': 139, 'Md.Pos': 56, 'Per': 29, 'Md.Nu/WQ': 5, Adv': 5, 'Md.Aj': 2, Dgcm.Br': 2, Dgcm': 1, App': 1, DP': 1, 'Mdcm.N': 1	'Md.Aj.Lx': 435, 'Md.Aj.Fn': 293, 'Md': 130, H': 83, 'Md.Nu/WQ.Nu': 40, Poss: 36, 'Md.Nu/WQ.WQ': 36, 'Dem': 34, 'GenP': 19, Q': 15, 'Md.Aj': 2, RC': 1, 'Mdmd': 1, Adv': 1	'Md.Aj.Lx': 62, 'Md.Aj.Fn': 21, Md': 8, Poss': 7, 'Md.Nu/WQ.Nu': 6, GenP': 5, 'Md.Nu/WQ.WQ': 3, Q': 2, Dem': 2, H': 1	'Md.Aj.Lx': 3, 'Md.Aj.Fn': 1, GenP': 1	
4825	1126	117	5	
5	117	1126	4825	N
Dem': 3, Q': 1, H': 1	Q': 7, 'Md.Nu/WQ.Nu': 6, 'Md.Aj.Fn': 4, 'Mdmd': 4, 'Per': 4, Poss: 3, 'Md.Pos': 1, 'Md': 1, 'Md.Aj.Lx': 1, 'Md.Nu/WQ.WQ': 10, H': 20, Dem': 56,	GenP': 6, Per': 4, Md': 4, Adv': 3, Dgcm.Br': 2, 'Md.Pos': 1, 'Mdcm.N': 1, Dgcm': 1, 'Md.Aj.Lx': 13, 'Poss': 35, 'Md.Nu/WQ.Nu': 37, 'Md.Aj.Fn': 39, 'Md.Nu/WQ.WQ': 39, Dem': 133, 'Mdmd': 135, Q': 137, H': 536,	'Md.Nu/WQ': 5, 'Md.Aj': 4, H': 4, Adv': 3, RC': 1, App': 1, 'Mdmd': 1, DP': 1, 'Per': 21, 'Md.Pos': 54, 'Md.Nu/WQ.WQ': 157, 'GenP': 202, Md': 285, 'Md.Nu/WQ.Nu': 394, Poss: 414, Q': 554, 'Md.Aj.Fn': 650, Dem': 824, 'Md.Aj.Lx': 1356,	

Average distance from N

- On average, some categories are closer to the noun, others further away
- The average slot-wise distribution, measured from both directions
processed and computed appropriately
will give us some abstract **value between 0 and** {NUMBER OF SLOTS}
indicating the respective **distance from the noun**

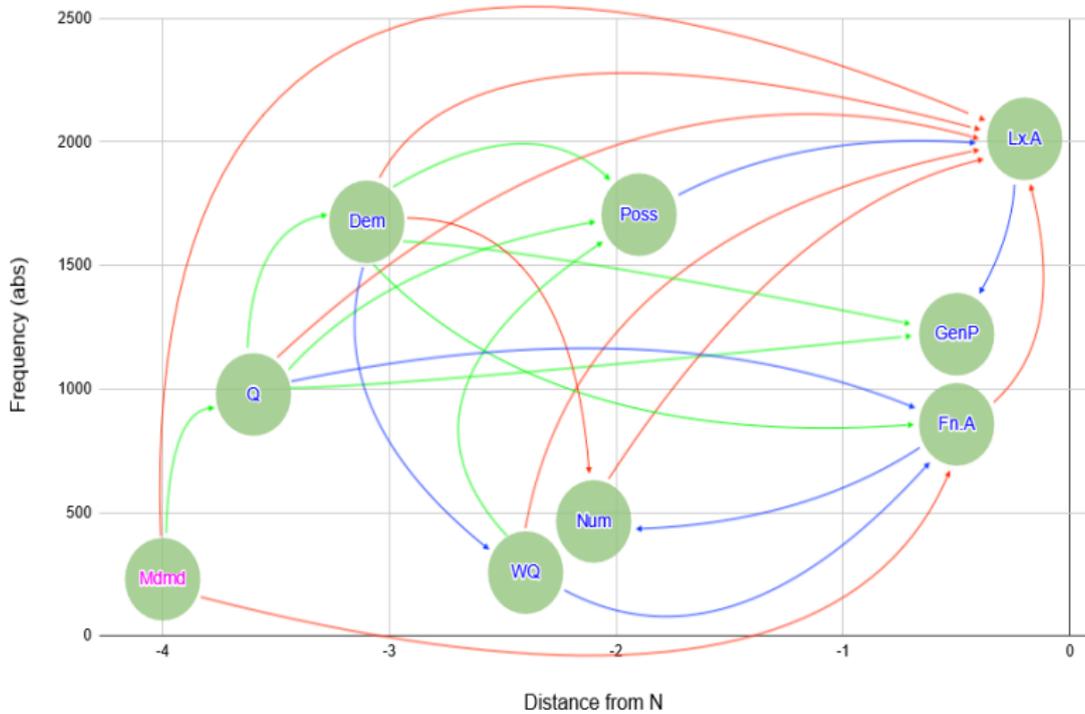
Method: **Probabilize**(level=2, x=False, part=-1, nom="N.C", mini=10)

Map the results on the Cartesian plane:

- prenominal cats = negative x-axis,
- postnominal cats = positive x-axis
- number of cat occurrences = (positive) y-axis

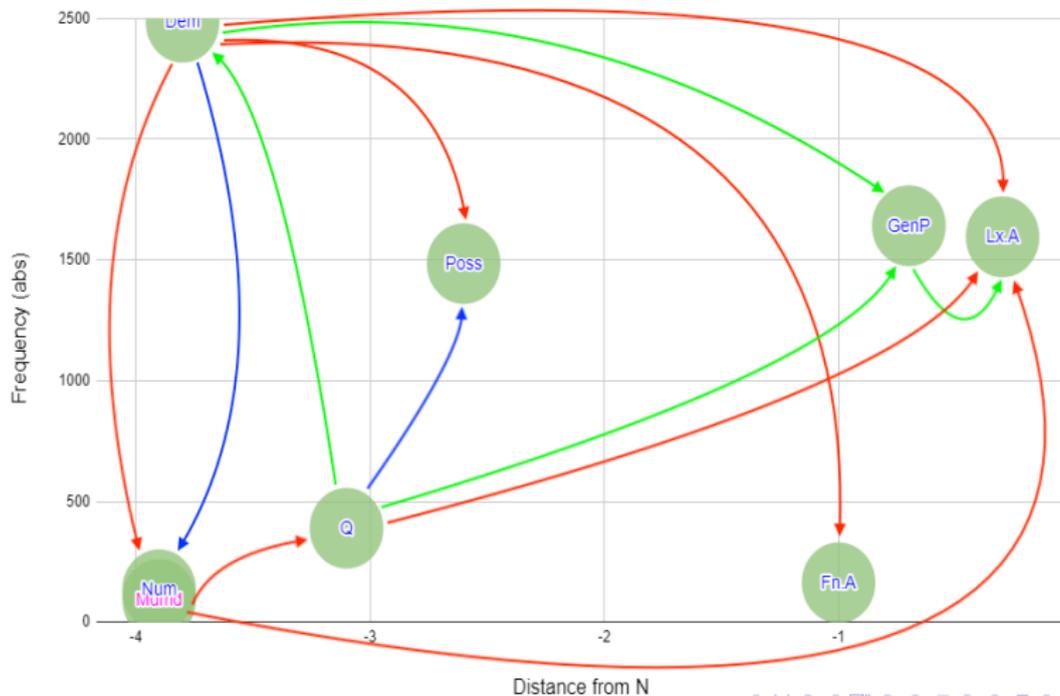
Schrödinger's CATs – Icelandic

Figure: Probabilistic categorial distribution in the Cartesian plane



Schrödinger's CATs – Old Saxon

Figure: Probabilistic categorial distribution in the Cartesian plane



Anything else to patternize?

- Besides NPs, basically any constituent can be patternized, e.g. VP
- Patternization of non-constituents: prenominal Domain, postnominal Domain
→ middle field
- Inclusion of other (morphological, semantic ...) features in pattern building
- Patterns ⇔ “Regular Expressions at a phrasal level?”
- Deducing a “Proto Pattern”?
- Incorporating other DBs/corpora / annotation styles

Takk fyrir mig ;)

Thank you very much for your attention!

- Pfaff, Alexander. 2014. Inside and outside – before and after. Weak and strong adjectives in Icelandic. In *Adjectives in Germanic and Romance*, eds. Petra Sleeman, Freek Van de Velde, and Harry Perridon, 217–244. Amsterdam: John Benjamins.
- Pfaff, Alexander. 2015. Adjectival and Genitival Modification in Definite Noun Phrases in Icelandic — A Tale of Outsiders and Inside Jobs. PhD diss, University of Tromsø.
- Pfaff, Alexander. 2016. *Icelandic articles and anaphora – the free article can be strong!* Ms., available at: <http://lingbuzz.auf.net/lingbuzz/003371>.
- Pfaff, Alexander. 2017. Adjectival inflection as diagnostic for structural position – Inside and outside the Icelandic definiteness domain. *Journal of Comparative Germanic Linguistics* 20 (3): 283–322.
- Pfaff, Alexander. 2019. Reunited after 1000 years. The development of definite articles in Icelandic. *Nordic Journal of Linguistics* 42 (2): 165–207.