Patternization – How to measure syntactic diversity?

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

May 4, 2024

Point of departure

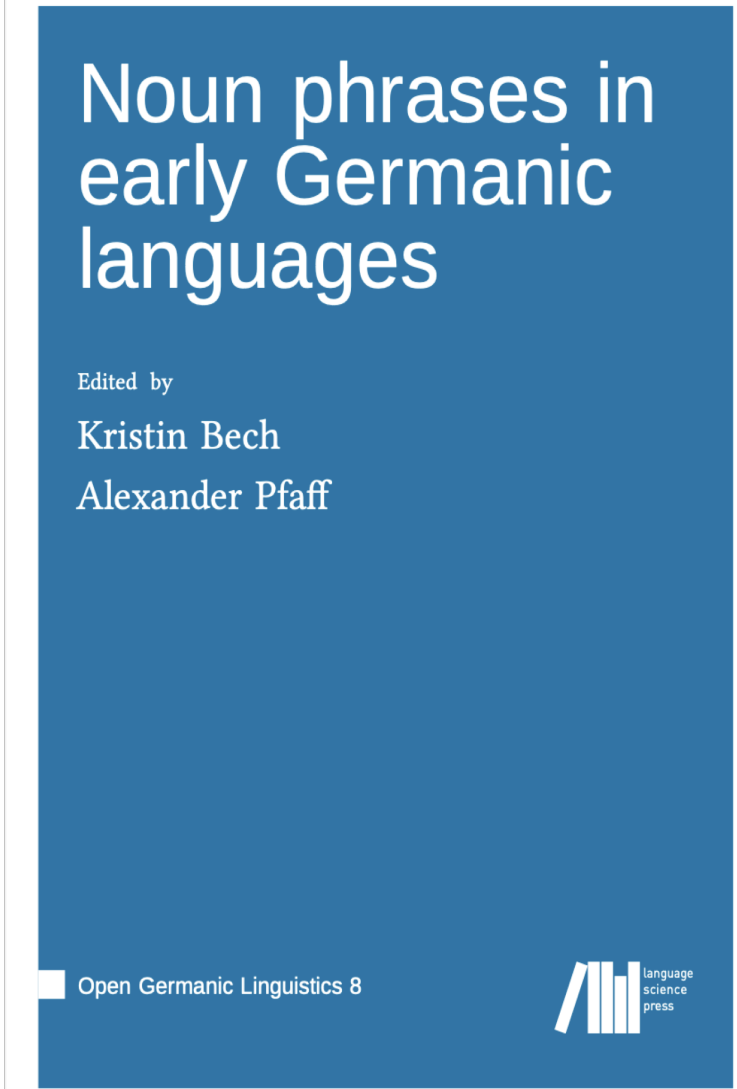
Constraints on Syntactic Variation:
***N**oun **P**hrases in **E**arly **G**ermanic **L**anguages** (= NPEGL)

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

- Research questions $\leftrightarrow$ nomen est omen
- corpora, quantitatve aproaches
- 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



Contents

https://langsci-press.org/catalog/book/436

Entry in the NPEGL database

LANGUAGE	Old Icelandic
DB ITEM ID	OIce.116.684
CONTEXT	Ekki hafði hann mikla ást af föður 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 föður 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 fróðleik 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 tungurnar eru ólíkar hver annarri, þær þegar er úr einni og hinni sömu tungu hafa gengist eða greinst, þá þarf ólíka stafi í að hafa en eigi hina sömu alla í öllum, sem eigi rita grikkir latínustöfum girskuna og eigi latínumenn girskum 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 [tungurnar]$_{tunga}$ {eru ólíkar hver annarri,} [þær]$_{sá}$ [þegar er úr einni og hinni sömu tungu hafa gengist eða greinst]$_{\#}$

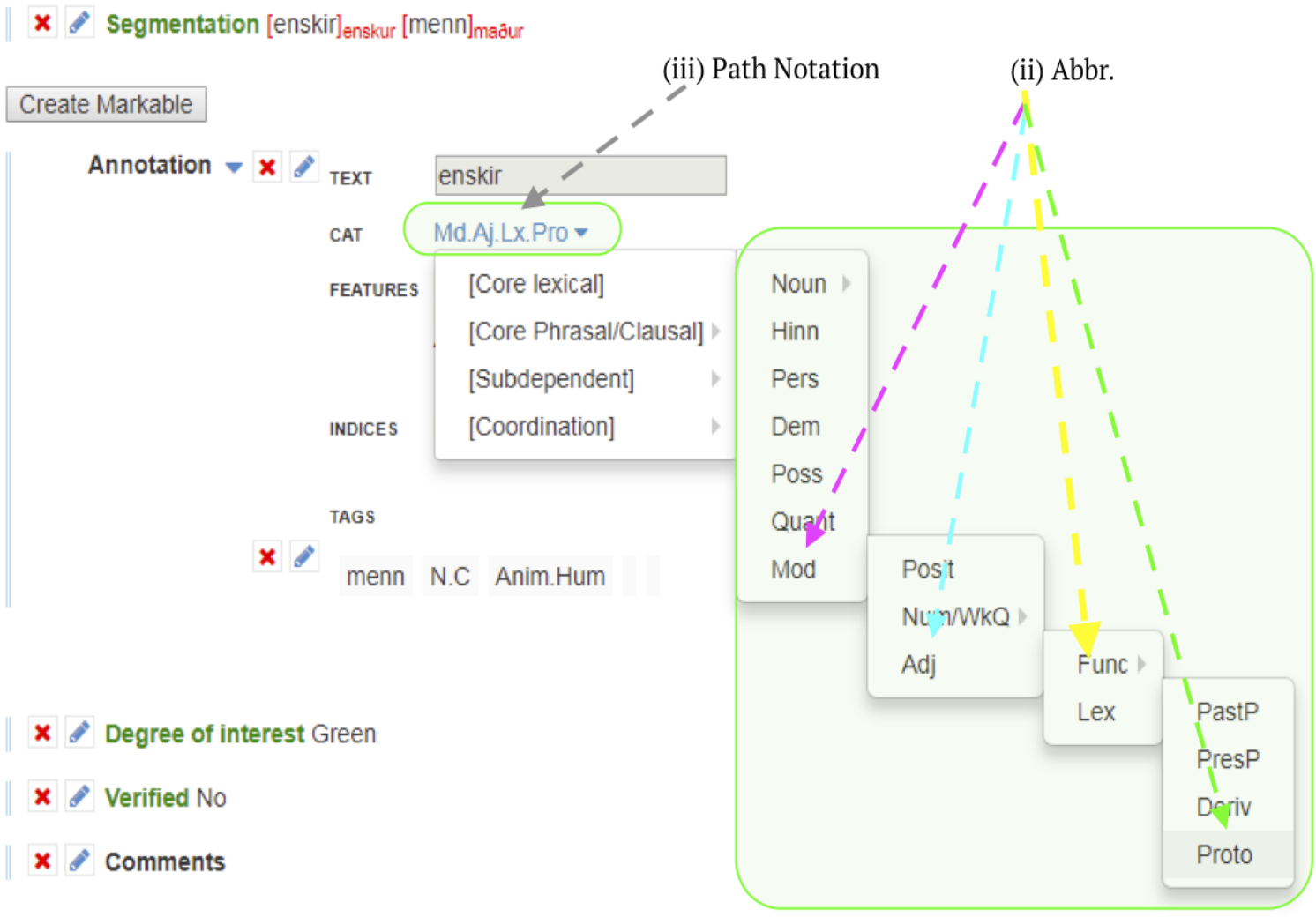
Create Markable

Annotation

tungurnar	N.C	Abst.Oth	Def, Sf
þær	Dem		Def
þegar er úr einni og hinni sömu tungu hafa gengist eða greinst	RC		

Degree of interest Green

The annotation interface: modifiers



The Modifier System

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

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" = *suffixed 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 fullkomn-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 fullkomn-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
(ii) N ADJ
(iii) POSS ADJ N
(iv) ADJ N POSS
(v) N POSS ADJ

(ix) Q ADJ N
(x) Q DET ADJ N
(x-α) Q N DET A
(xi) ADJ POSS N
(xii) Q N

(xv) POSS N RC
(xvi) N DEM POSS ADJ

. . . . . .

(???-a) . . . . . .

Patterns: Linear sequences of Cat labels

	þessir	*tveir*	*stóru*	*hestar*	
cat^0	**Dem**	**Md**	**Md**	**N**	$\rightarrow$ ***patt***^{0}
cat^1	Dem	**Md.Card**	**Md.Aj**	**N.C**	$\rightarrow$ ***patt***^{1}
cat^2	Dem	**Md.Card.Nu**	**Md.Aj.Lx**	N.C	$\rightarrow$ ***patt***^{2}
cat^3	Dem	Md.Card.Nu	**Md.Aj.Lx.Pro**	N.C	$\rightarrow$ ***patt***^{3}

Table: *Pattern construal @ 4 levels of the NP "these two big horses"*

patt^0 : (Dem, Md, Md, N)
patt^1 : (Dem, Md.Card, Md.Aj, N.C)
patt^2 : (Dem, Md.Card.Nu, Md.Aj.Lx, N.C)
patt^3 : (Dem, Md.Card.Nu, Md.Aj.Lx.Pro, N.C)

patterns are encoded as tuples

mixed : (Dem, Md, Md.Aj.Lx, N.C) $\rightarrow$ 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) $\mu\text{-}p = \frac{1}{n} \sum_{i=1}^{n} p_i \quad \Rightarrow \quad \textsc{PattDiv} = \frac{\mu\text{-}p}{scd}$

(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 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

(3)

⇒	{N, POSS, ADJ }		
i.	**POSS ADJ N**	[+ATT]	**1 / True**
ii.	**ADJ N POSS**	[+ATT]	**1 / True**
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 x (n-1) x (n - 2) x ... x 2 x 1 (full) ***permutations*** (of n)
- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ ***k-combinations*** (of n) $\Rightarrow$ sub-sets of size k
- $\binom{n}{k}$ x $k!$ = $\frac{n!}{(n-k)!}$ ***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 $\binom{6}{3} = \frac{6!}{3!(6-3)!}$ = **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 x** 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) . . . $\neq$ {A, B, C}

Relevant sample spaces & how to use them

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

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

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

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

S_{cat} = $cat^2 \cup$ {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: `bool`ean (= attestation: yes/no). or `int`eger (= 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-earch and M-atch Patterns

⇒ **S-Patterns**
→ F(*patt*)

M-Patterns ⇐
match/result ←

- `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 |**

Some results in the Old Icelandic ndb

(4)

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

	OldIcel	**OldEngl**	**OldHighGer**	**OldSwed**	**OldSax**
categories	28	27	23	24	23
$\mathbf{C}_{pot}$	351	325	231	253	231
$\mathbf{C}_{att}$	153	99	52	58	81
$\mathbf{R}^{att}_{pot} = \frac{C_{att}}{C_{pot}}$	0.436	0.305	0.225	0.229	0.351
μ-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*

[1] For instance, Old Icelandic has 28 ATTESTED cat[2] 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}

∀ cat ∈ CATs:
∀ np ∈ ndb:

- → `Right_flexi_pattern(np, "N.C", cat, "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] → 52

- → `flexi_pattern(np, "N.C", cat, "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 . . .]

- → `flexi_pattern(np, "N.C", cat, "Dem", "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** . . .]

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

Combinatorial Flexibility: { N.C, Md.Aj, **X** } — for X = {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.	(Q, Md.Aj, N.C):	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.	(Q, N.C, Md.Aj):	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.	(Md.Aj, N.C, Q):	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.	(N.C, Q, Md.Aj):	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.	(Md.Aj, Q, N.C):	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.	(N.C, Md.Aj, Q):	1	02	2
		2	*006*	*1.6*

Co-Distribution

(6)	a.	(**x**, Md.Aj, N.C): 5	x ∈ {WQ, Dem, Num, Poss, Q}
	b.	(**x**, N.C, Md.Aj): 5	x ∈ {WQ, Dem, Num, Poss, Q}
	c.	(Md.Aj, **x**, N.C): 5	x ∈ {WQ, Dem, Num, Poss, Q}
	d.	(Md.Aj, N.C, **x**): 4	x ∈ {Dem, Num, Poss, Q}
	e.	(N.C, **x**, Md.Aj): 4	x ∈ {Dem, Num, Poss, Q}
	f.	(N.C, Md.Aj, **x**): 1	x ∈ {Q}

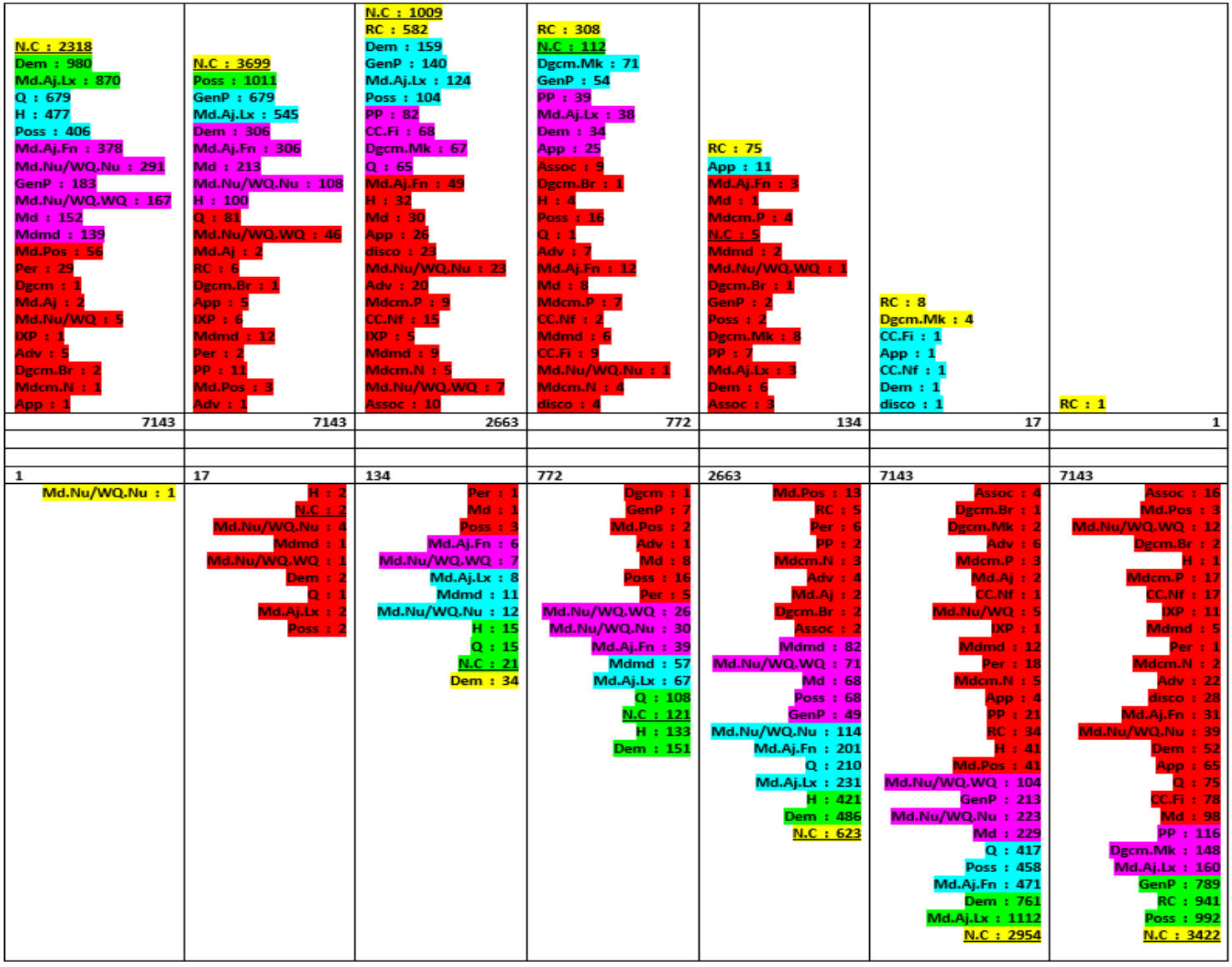
⇒ Dem, Poss and Num have an identical distribution in the context of N & A:
– all three occur in the permutations (6a-e),
– all three do not occur in the permutation (6f)

Aligning & stacking NPs

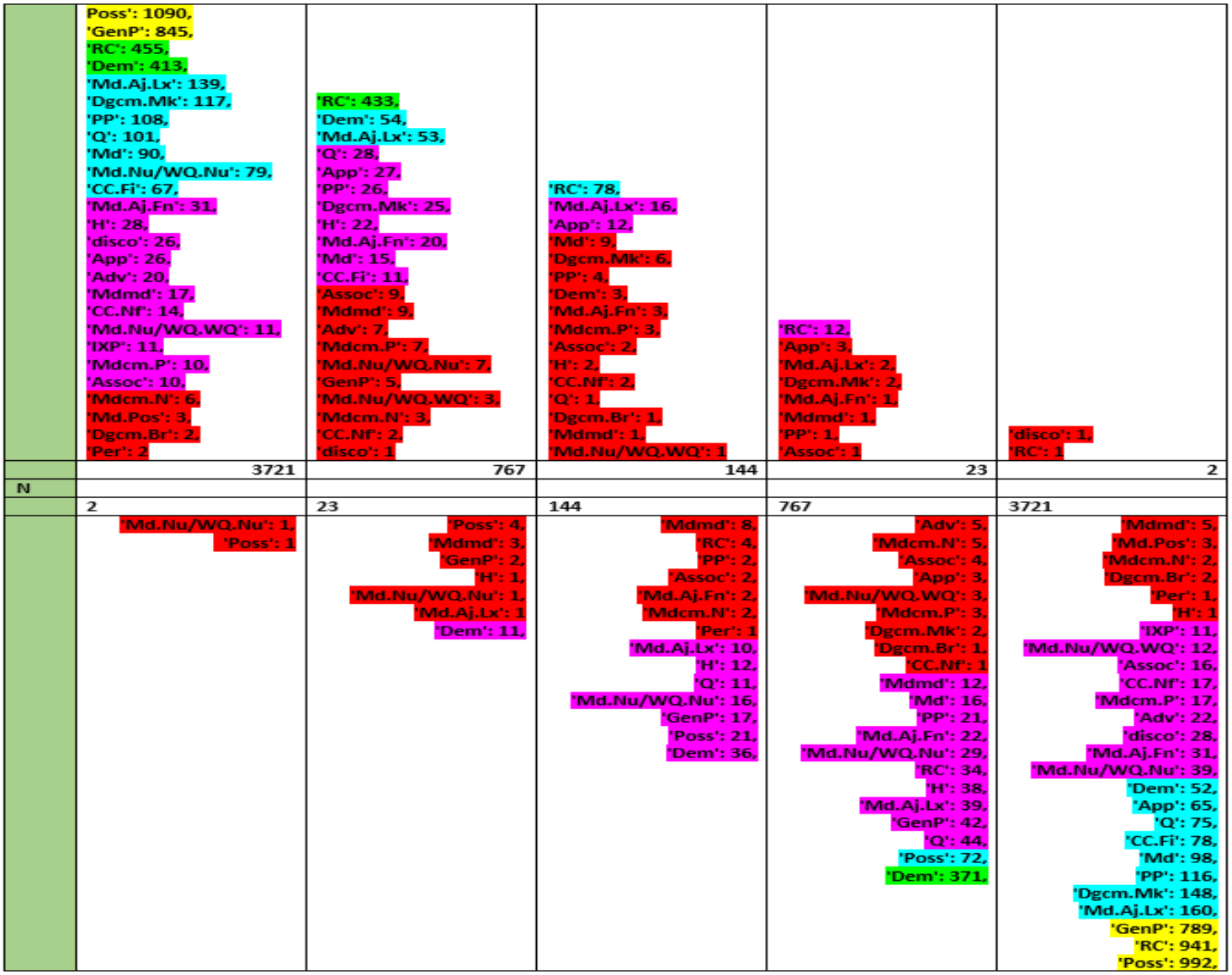
1	2	3	4	5
Dem	Adj	N		
Adj	N			
Dem	N	RC		
Q	Adj	N	Dem	RC
N	Adj	Dem	RC	
Dem	Adj	N		
Adj	N	GenP		
Num	Adj	N	GenP	
Dem	Num	N		
Dem	N	RC		
. . .	. . .	. . .	. . .	. . .

-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
. . .	. . .	. . .	. . .	. . .

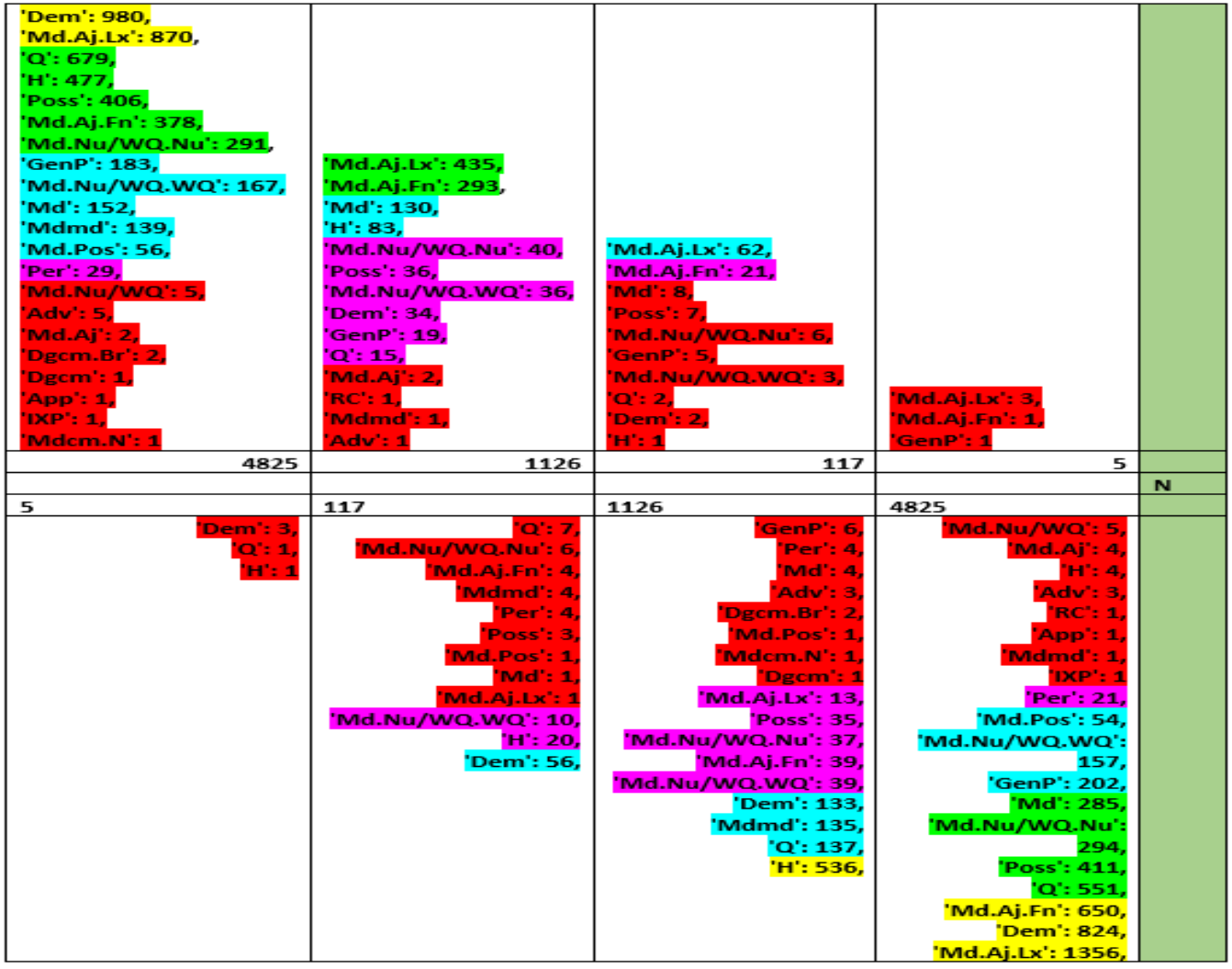
The Proto NP – a probabilistic model



The Proto NP – a probabilistic model; postnominal field



The Proto NP – a probabilistic model; prenominal field



Average distance from N

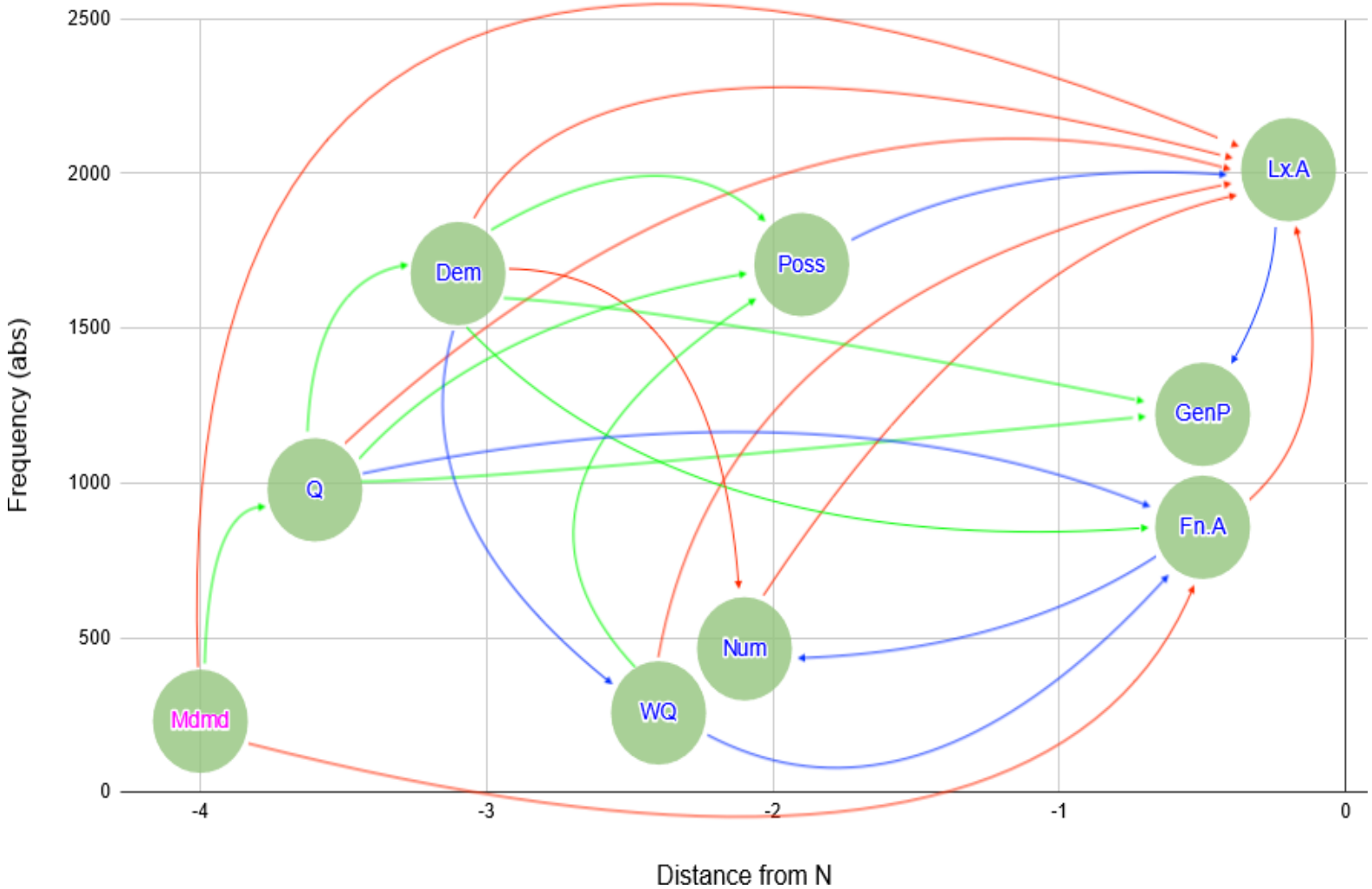
- On average, some categories are closer to the noun, others further away
- The average slot-wise distribution, measured from both directions
  processed and computed appropriately
  will give us some abstract **value between 0 and** {NUMBER OF SLOTS}
  indicating the respective **distance from the noun**

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

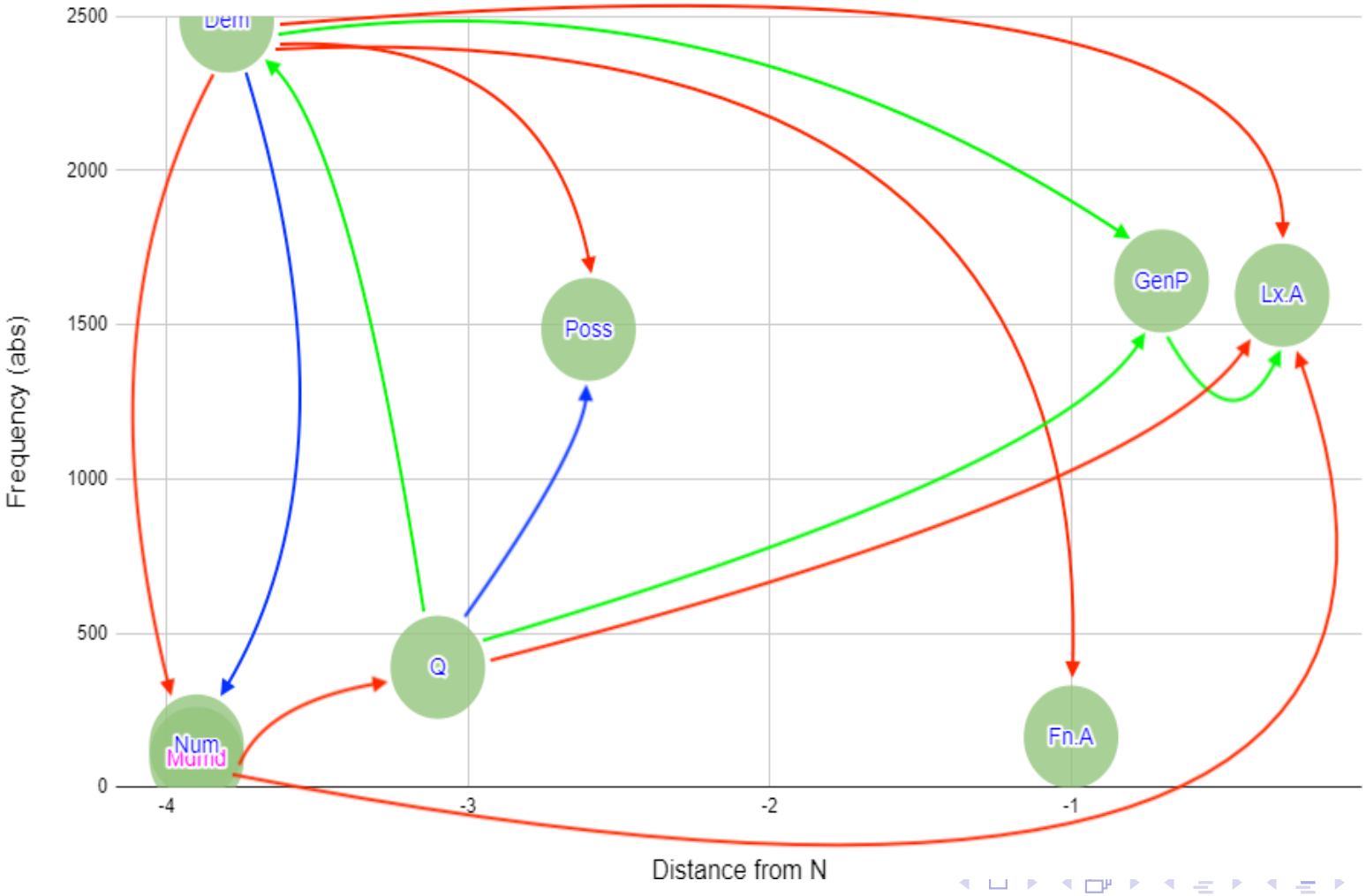
Schrödingers CATs – Icelandic

Figure: *Probabilistic categorial distribution in the Cartesian plane*



Schrödingers 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 ....
  $\rightarrow$ middle field
- Inclusion of other (morphological, semantic ...) features in pattern building
- Patterns $\Leftrightarrow$ “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.