

P17-0716:1

Swedish Grammar Profile

Therese Lindström Tiedemann* (therese.lindstromtiedemann@helsinki.fi)

Based on work done with: Elena Volodina¹, Yousuf Ali Mohammed*¹, Anders Agebjörn¹, Maisa Lauriala*

*=University of Helsinki, ¹= University of Gothenburg

Workshop on profiling L2 vocabulary & grammar, Gothenburg, Sweden (20 April 2023)



UNIVERSITY OF GOTHENBURG

SPRÅKBANKENTEXT



UNIVERSITY OF HELSINKI

<https://spraakbanken.gu.se/larkalabb/svlp>
(username: demo)



“L2profiles”

Development of Lexical and
Grammatical Competences in
Immigrant Swedish



UNIVERSITY OF GOTHENBURG

SPRÅKBANKEN **TEXT**



Project financed by

Riksbankens jubileumsfond (2018–2020 (2022))
P17-0716:1

<https://rj.se/anslag/2017/development-of-lexical-and-grammatical-competences-in-immigrant-swedish/> (incl. brief report)

Collaboration:

Swedish language, University of Gothenburg
Dept of Finnish, Finno-ugrian and Scandinavian
Studies, University of Helsinki

PI: Elena Volodina, Gothenburg

Homepage:

<https://spraakbanken.gu.se/en/projects/l2profiles>



UNIVERSITY OF HELSINKI

Today

- Aims
- Methods
- Examples / demo
- Throughout: future work & lessons learnt
- Summary

Our wishlist for grammar profiling...

- Inflections, parts of speech...
- Word order development, e.g.
 - What occurs in the first position (Sw. fundamentet)?
 - Is the finite verb always in the second position?
 - Is the clause adverbial in the correct position (depends on the type of clause)?
- Noun phrases –
 - Simple: only the head noun or a pronoun? OR
 - Extended: if so with which type of modifiers?
 - Definite or indefinite? Normlike usage?
- Verb phrases –
 - Which tenses?
 - Which type of passive?
 - Are simple or periphrastic forms learnt first and which is more challenging?
 - When do learners understand which past tense / passive to use?

Which can we tackle best through the annotation?

→ how well does the automatic annotation work?

Our wishlist

- Inflections, parts of speech...
- Word order development, e.g.
 - What occurs in the first position (Sw. fundamentet)?
 - Is the finite verb always in the second position?
 - Is the clause adverbial in the correct position (depends on the type of clause)?
- Noun phrases –
 - Simple: only the head noun or a pronoun? OR
 - Extended: if so with which type of modifiers?
 - Definite or indefinite? Normlike usage?
- Verb phrases –
 - Which tenses?
 - Which type of passive?
 - Are simple or periphrastic forms learnt first and which is more challenging?
 - When do learners understand which past tense to use?

How reliable is the automatic linguistic annotation?

- POS & lemmatisation (incl. POS, lemgram) work quite well on course books, learner essays and normalised learner essays.
 - Dependency analysis not very good.
- We **cannot** rely on the dependency analysis
- Queries must primarily use: POS; MSD; lemgrams.

Volodina, E., Alfter, D., Lindström Tiedemann, T., Lauriala, M. S., & Piipponen, D. H. (2022). Reliability of automatic linguistic annotation: native vs non-native texts. In *Selected papers from the CLARIN Annual Conference 2021*. Linköping University Electronic Press (LiU E-Press).

Receptive and productive

- Learning a language
 - learning to understand (RECEPTIVE SKILLS)
 - learning to produce (PRODUCTIVE SKILLS)

Receptive and productive

- Learning a language
 - learning to understand (RECEPTIVE SKILLS)
 - learning to produce (PRODUCTIVE SKILLS)

→ Two corpora:

- L2 Course book corpus (*Coctail*, Volodina et al 2014) – RECEPTIVE
- L2 learner corpus (*SweLL-pilot*, Volodina et al 2016) – PRODUCTIVE
- Both linked to CEFR (COE, 2001)

→ Compare the development in receptive and productive proficiency.

Usage-based perspective – input important

- Learning is partly implicit.
 - Frequency affects learning (cf e.g. Bybee 2008)
- Important to see how **course books** and **learner essay** usage compares.

Comparison to L1 usage

- Important to be able to compare to reference corpora
 - Open queries which can be rerun on reference corpora

Interlanguage use and norms

- No error/correction annotation
- **Normlike patterns**
(based on the most authoritative descriptive grammar of Swedish, the Swedish Academy Grammar (Teleman et al. 1999))
- Interlanguage use of normlike patterns – overviews in the profile not showing whether correctly used
- Profile → open data → such analysis possible



Language Acquisition Reusing **Korp**

Lexical profile

Grammatical profile

Morphological profile

<https://spraakbanken.gu.se/larkalabb/svlp> (username: demo)

Morphology:
word formation patterns

Noun patterns

Verb patterns

Noun patterns



Language Acquisition Reusing **Korp**

Lexical profile

Grammatical profile

Morphological profile

Adjectival declension

Adjectival & Adverbial structure

Multi Word Expressions

Sen*Lex

Adjectiva

Sen*lex can be filtered by:

- gender
- nominal declination
- verbal conjugation

Grammar profile

- grammatical features in the lexical profile



Language Acquisition Reusing **Korp**

<https://spraakbanken.gu.se/lar/kalabb/svlp> (username: demo)

Morphology:
word formation patterns

Lexical profile

Grammatical profile

Morphological profile

Verb patterns

Noun patterns



Language Acquisition Reusing **Korp**

Lexical profile

Grammatical profile

Morphological profile

Adjectival declension

Adjectival & Adverbial structure

Multi Word Expressions

Sen*Lex

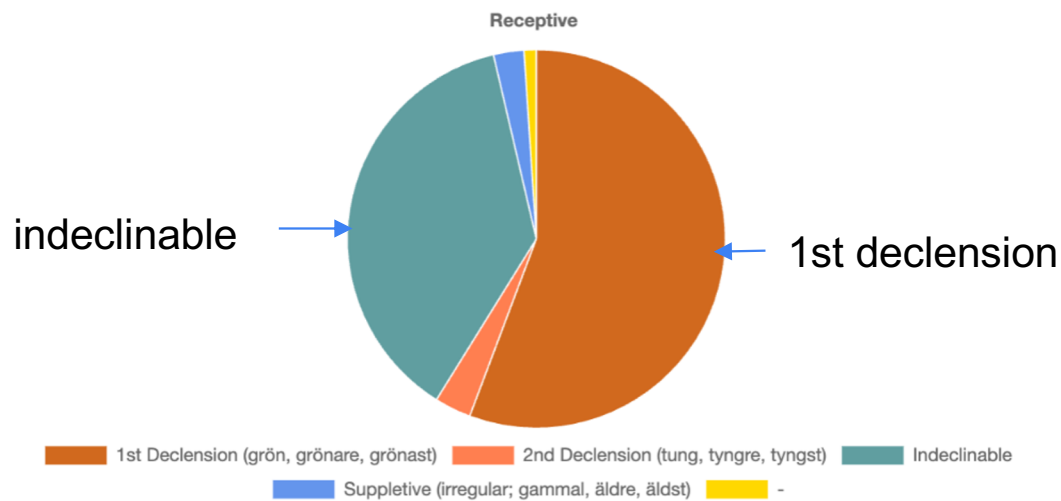
Adjectiva

Sen*lex can be filtered by:

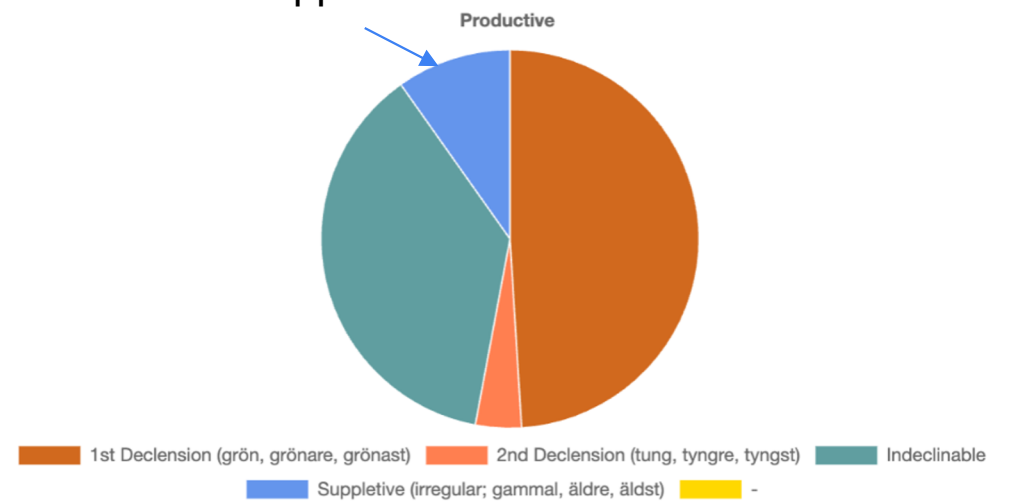
- gender
- nominal declination
- verbal conjugation

Adjectival declension (A1)

Adjectival declension

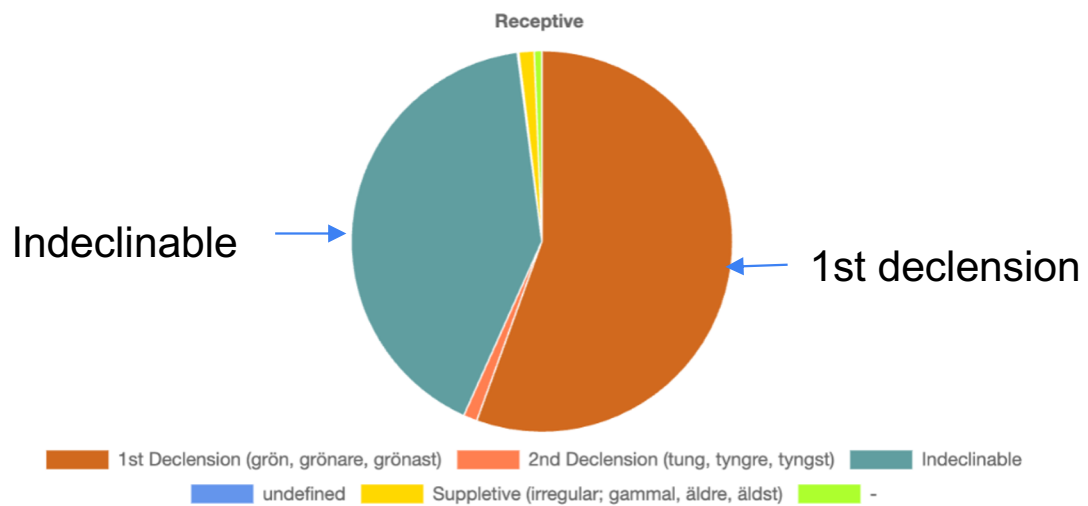


suppletive

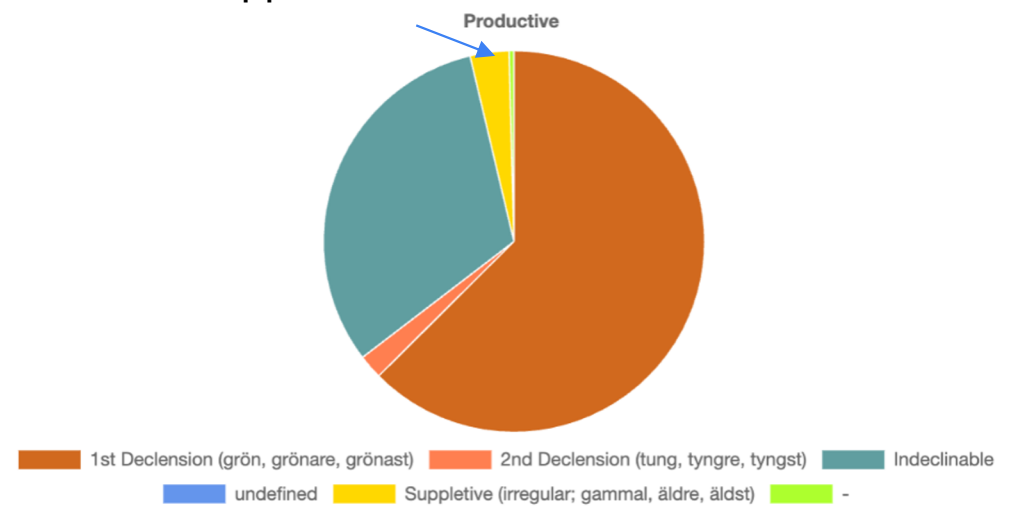


Adjectival declension (B1)

Adjectival declension

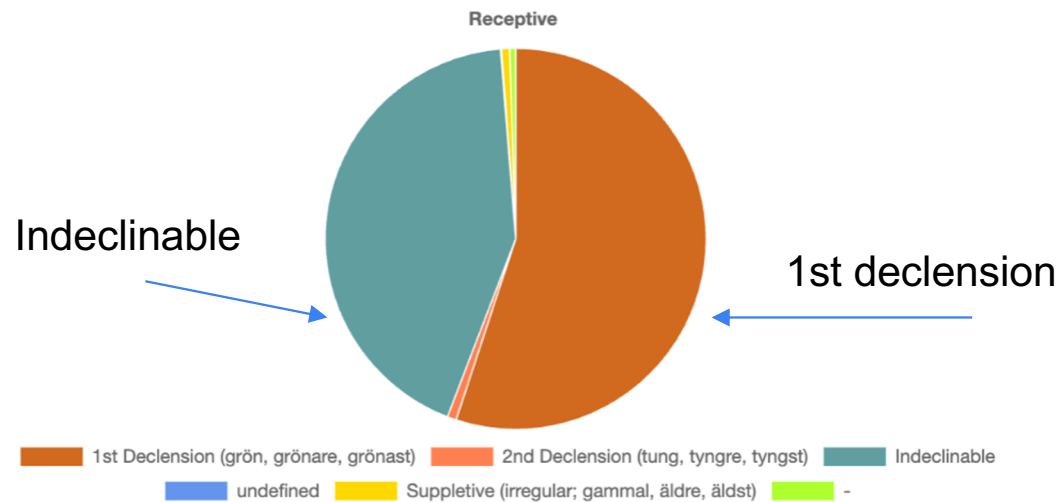


suppletive

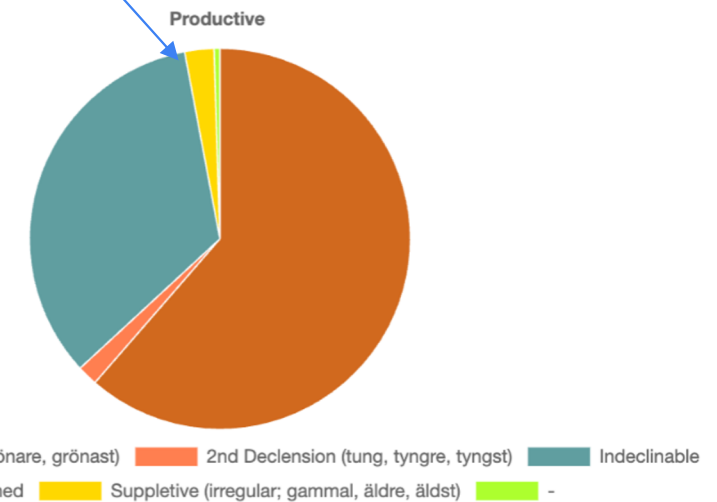


Adjectival declension (C1)

Adjectival declension



Suppletive



- Learners use suppletive adjectives more than they see them in texts.
- Lemma-based → Do they mainly use the positive form, or do they also also use the suppletive comparative and superlative forms?
- If they do, are they used correctly?

Which forms are used?

Adjectival declension [🔗](#)

Adjectival declension
Suppletive (irregular; gammal, äldre, äl... [▼](#))

Saldo Word Class [▼](#)

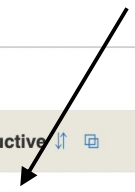
A1 A2 B1 B2 C1 C2

Tables Graphs Statistics Download Clear all

Show only first occurrence [Tables - description](#) [Filters - description](#)

Total rows : 50

CEFR level ↕	Lemgram ↕	Sense ↕	Word Class ↕	Saldo Word Class ↕	Adjectival declension ↕	Receptive ↕ 🔗	Productive ↕ 🔗
B2	liten..av.1	liten..1	Adjective (JJ)	Adjective (av)	Suppletive (irregular; gammal, äldre, äldst)	12.36 (187)	5.17 (23)
B1	bra..av.2	bra..4	Adjective (JJ)	Adjective (av)	Suppletive (irregular; gammal, äldre, äldst)	13.91 (184)	28.04 (64)
B1	gammal..av.1	gammal..1	Adjective (JJ)	Adjective (av)	Suppletive (irregular; gammal, äldre, äldst)	13.76 (182)	10.95 (25)
B1	liten..av.1	liten..1	Adjective (JJ)	Adjective (av)	Suppletive (irregular; gammal, äldre, äldst)	13.31 (176)	10.52 (24)



Statistics – compile by MSD

KWIC: hits per page: 25 ▾ sort within corpora: not sorted ▾ Statistics: compile based on: word (+1) ▾ Show statistics Show word picture

KWIC Statistik Ordbild

Results: 25

« < 1 > » Go to page of 1 [Show context](#)

word **Aa**

Word attributes

baseform

part-of-speech

msd

lemgram

dependency relation

SPIN (does not support extended context)

Hon var född i en **liter** **lay** i Dalsland.

B2 level (prod.) mainly the positive (7+6+2+1=16),
comparitive (8)

The positive form needs to agree with
the main word.

Singular and plural agreement also
show suppletion.

*En **liten** artikel 'a small article'*

***Små** artiklar 'small articles'*

KWIC Statistik Ordbild

Show Trend Diagram

Show map

Number of rows: 7

<input type="checkbox"/>	word	msd	Total
<input checked="" type="checkbox"/>	Σ	Σ	159.2 (25)
<input type="checkbox"/>	mindre	JJ.KOM.UTR+NEU.SIN+PLU.IND+...	51 (8)
<input type="checkbox"/>	liten	JJ.POS.UTR.SIN.IND.NOM	44.6 (7)
<input type="checkbox"/>	små	JJ.POS.UTR+NEU.PLU.IND+DEF....	38.2 (6)
<input type="checkbox"/>	lilla	JJ.POS.UTR+NEU.SIN.DEF.NOM	12.7 (2)
<input type="checkbox"/>	minst	JJ.SUV.UTR+NEU.SIN+PLU.IND.N...	6.4 (1)
<input type="checkbox"/>	litet	JJ.POS.NEU.SIN.IND.NOM	6.4 (1)

Grammatical features for lexical items

The screenshot displays the Sen*Lex interface. At the top, three tabs are visible: 'Lexical profile' (highlighted in dark teal), 'Grammatical profile' (light beige), and 'Morphological profile' (light beige). Below the tabs, the 'Sen*Lex' logo is on the left. A horizontal line separates the header from the search area. The search area contains a 'Search word' input field and five dropdown menus: 'Word Class', 'Saldo Word Class', 'Noun declension', 'Gender', and 'Conjugation'. Below the search area is a row of six colored buttons: A1 (dark grey), A2 (green), B1 (light blue), B2 (dark blue), C1 (brown), and C2 (teal). To the right of these buttons are five more buttons: 'Tables' (dark teal), 'Graphs' (light blue), 'Statistics' (light blue), 'Download' (gold), and 'Clear all' (gold). At the bottom left, there is a checkbox labeled 'Show only first occurrence' and two links: 'Tables - description' and 'Filters - description', both with external link icons.

Manual checks of lists necessary.

Partly populated from other resources at Språkbanken

- POS-tagging
- POS as part of lemgram (Saldo word class)
- Noun declension
- Gender
- Conjugation

Most common adjectives at A1

Lexical profile – Sen*lex

Saldo word class

- Adjective
- Adjective abbreviations
- Multiword adjectives

Receptive (course books)

- Halv 'half'
- **Liten 'small'**
- **Stor 'big'**
- Ny 'new'
- Svensk 'Swedish'

Productive (learner essays)

- **Bra 'good'**
- Halv 'half'
- Svensk 'Swedish'
- Trött 'tired'
- Somalisk 'Somalian'

Parts of speech, e.g. prepositions

Lexical profile Grammatical profile Morphological profile

Two ways of approaching word classes –
POS-tagging and lemmgrams from the **Saldo**
lemmatisation

Sen*Lex 

Word Class 

A1 A2 B1 B2 C1 C2

Show only first occurrence [Tables - description](#)  [Filters - des](#)

Total rows : 309

Saldo Word Class

- Prefix (sxc)
- Multiword prefix (mxc)
- Preposition (pp)
- Multiword preposition (ppm)
- Preposition, abbreviation (ppa)
- Pronoun (pn)

Prepositions (Saldo lemgrams)

	A1	A2	B1	B2	C1
Receptive	26	36	64	78	77
Productive	7	21	32	40	47
Productive prepositions	I, på, till, från, utan, av...	I, på, till, från, av, hos...	I, på, till, av, från, utan...	I, på, av, till, från, på grund av...	I, på, av, till, på grund av, från...
Eng. translations	In, on, to, from, without, of	In, on, to, from, of, at (someone's house)	In, on, to, of, from, without	In, on, of, to, from, because...	In, on, of, to, because, from



Language Acquisition Reusing [Korp](#)

Lexical profile

Grammatical profile

Morphological profile

Noun patterns

Verb patterns

Noun patterns

Norm-based patterns

- All patterns are based on Swedish L1 norms (based on SAG, the Swedish Academy Grammar, Teleman et al. 1999 – **descriptive grammar**)
- Important to use the most widely accepted description of the language as a foundation.
 - Research-based → possible to use for both research and teaching

Norm-based patterns

- Non-normlike PATTERNS will not usually be found,
*e.g. *en boken 'a book-DEF'*
- Non-normlike USAGE might still be found – a pattern might be used in the wrong context...,
*e.g. *Jag spelar gitarren. 'I play-PRES. guitar-DEF.'*
- **Possible future work:** non-norm patterns
(some already available in the documentation but not included in the profiling tool)

Best overviews → Only varying clause position



Noun patterns 

Pattern (description) ▼ Pattern (structure) ▼ Clause position
Varying ▼ Definiteness ▼

Gender ▼ Number ▼ Attribute ▼ Article ▼ Other definite attributes ▼

Initial clause position –

makes it easier to capture some patterns.

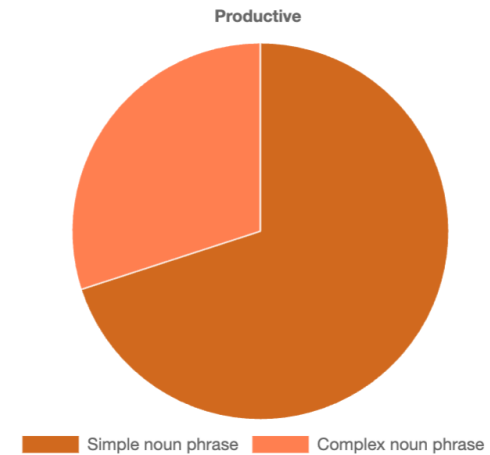
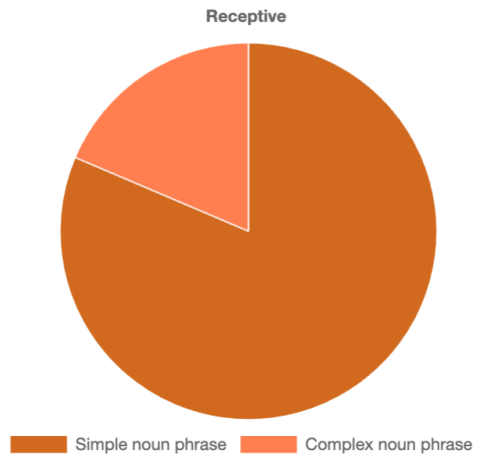
Also included in the varying clause position. →
important to exclude from overviews.

(→ **default setting?**)

A1 more complex NPs in production...

- Varying clause position
- All occurrences

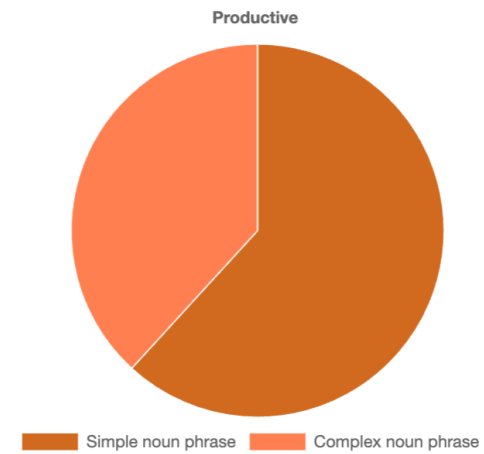
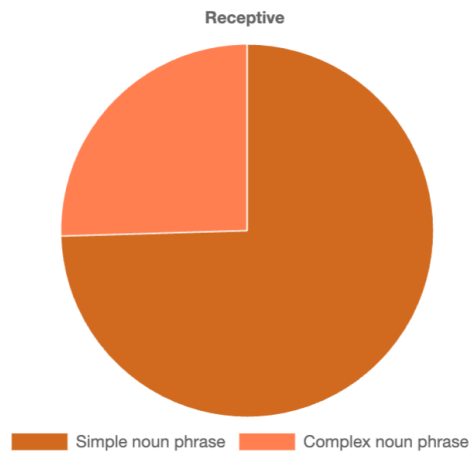
Pattern (structure)



A2 more complex NPs in production...

- Varying clause position
- All occurrences

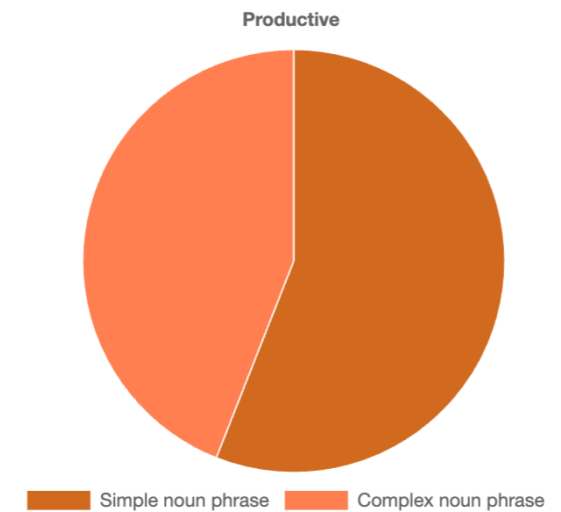
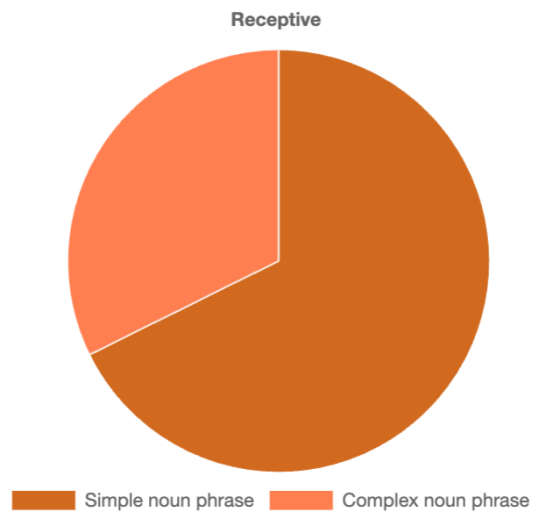
Pattern (structure)



B1 more complex NPs in production...

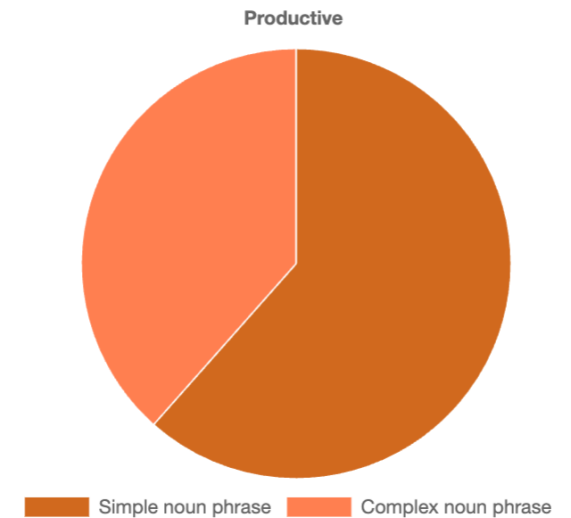
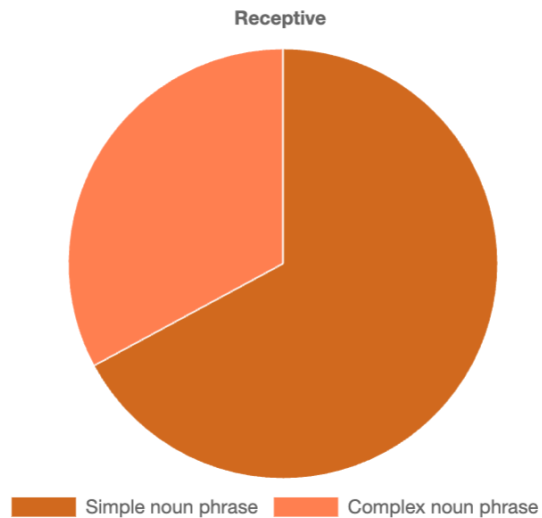
- Varying clause position
- All occurrences

Pattern (structure)



B2 more complex NPs in production... – but nearly equal

Pattern (structure)



- Varying clause position
- All occurrences

C1 more or less equal...

- Varying clause position
- All occurrences

Pattern (structure)



- Assessment – complex NPs are seen as part of higher proficiency
- Are complex NPs used in the same way in learner writing as in L1?
- Is there overuse? Topic related? Assessment driven?

Most common complex NPs (varying position) – A1 prod.

15c. PossINP-gen Nsg (<i>min kompis</i>)	Complex noun phrase	1.65 (125)	16.75 (106)
6e. A Npl (kör <i>gula bilar</i>)	Complex noun phrase	2.06 (156)	7.27 (46)
2e. A Nsg (när hon var <i>liten flicka</i>)	Complex noun phrase	1.54 (117)	5.69 (36)
6f. A Npl (att använda <i>alternativa sätt</i>)	Complex noun phrase	0.91 (69)	5.06 (32)
15d. PossINP-gen Nsg (<i>lärarens telefonnummer</i>)	Complex noun phrase	0.74 (56)	2.05 (13)
2f. A Nsg (med <i>stort intresse</i>)	Complex noun phrase	0.42 (32)	1.26 (8)

2e A Nsg:

”när hon var **liten flicka**”
 when she was **little girl**
 ’when she was **a little girl**’

’Naked nouns’

Can only be used in certain contexts.
 There may be errors in usage.

Language data

Quite a lot of language examples:
Jag talar svenska. 'I speak Swedish'

Results: 36

« < 1 2 > » Go to page of 2 [Show context](#)

SPIN V1 (does not support extended context)

I skolan är	inte bra datore	.
Jag talar somaliska	lite Engelska och lite svenska	.
Jag talor Tigirigina	och lite svaneska	.
Somalia	har tva åstid	vinten vår, sommar och höst.
Jag studera	svensk etar\$ d matmmatik.	.
Jag kommer Från Somalia och	talar somaliska och linte svenska	.
kan half nio I ella dagen också vi slutar skolan till example m\$ \$ndag vi slutar kockan två tisdag klockan tre over tjugo andra dag vi	sluta diffrent timma	.
Han	är mycket bra lärare	.
Han är gift Eleverna är 16 i min klassrumet Alla eleverna är 350 i skolan Jag drincker Cofe Efter rast Vi har en lektion Den skolan	är mycket bra skola	Jag spelar fotball Min skolan börjar 8:30 slutar 11:30 Ef
personer arbetar i skolan är lärare sjukhuset Bra Lärare, bra elever	och bra klass\$ Idrott	, Matematik Svenska.
Jag börj\$ r iskolan till December, sluten skolan nest decemb Jag	åker tilliskolan buss	.
Jag	talar arabiska och liten svaniska	.
	Jag alskare vår	.

Counter-norm usage:

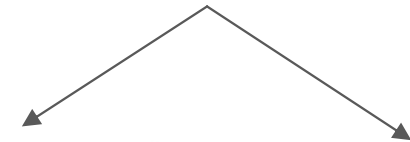
*Den skolan är __ mycket **bra skola**.
Den skolan är **en** mycket **bra skola**.
That school is a very **good school**.

Compare to a learner corpus with correction labels... SweLL gold

Elena Volodina, Lena Granstedt, Arild Matsson, Beáta Megyesi, Ildikó Pilán, Julia Prentice, Dan Rosén, Lisa Rudebeck, Carl-Johan Schenström, Gunlög Sundberg and Mats Wirén (2019). The SweLL Language Learner Corpus: From Design to Annotation. Northern European Journal of Language Technology, Special Issue.

Frequency – link to search → reference corpora

Click on the frequency → Hits in Korp & query → can change the corpus



CEFR level ↕	Pattern (description) ↕	Pattern (structure) ↕	Receptive ↕ 📄	Productive ↕ 📄
A1	1b. Nsg (Jag kan <i>svenska</i>)	Simple noun phrase	1.00 (76)	1.42 (9)
A2	1b. Nsg (Jag kan <i>svenska</i>)	Simple noun phrase	1.04 (145)	1.41 (37)
B1	1b. Nsg (Jag kan <i>svenska</i>)	Simple noun phrase	0.74 (137)	1.57 (27)
B2	1b. Nsg (Jag kan <i>svenska</i>)	Simple noun phrase	0.83 (139)	2.75 (74)
C1	1b. Nsg (Jag kan <i>svenska</i>)	Simple noun phrase	0.89 (148)	3.26 (94)
C2	1b. Nsg (Jag kan <i>svenska</i>)	Simple noun phrase	NaN (0)	2.86 (5)

Change corpus → e.g. SweLL

- Remove the level from the search

KORP v9

2 of 275 corpora selected — 652.00K of 15.19G tokens

SPRÅKBANKENTEXT SWE-CLARIN

Simple Extended **Advanced** Compare

Active CQP query in simple search:

Active CQP query in extended:
[]

Custom CQP query:
[Get the CQP tutorial](#)
[(msd = "NN\,UTR\,SIN\,IND\,NOM") & pos = "NN" & lex not contains "januari\.\.nn\,1" & lex not contains "februari\.\.nn\,1" & lex not contains "mars\.\.nn\,1" & lex not contains "april\.\.nn\,1" & lex not contains "maj\.\.nn\,1" & lex not contains "juni\.\.nn\,1" & lex not contains "juli\.\.nn\,1" & lex not contains "augusti\.\.nn\,1" & lex not contains

Remove the level

















SweLL gold original – correction labels & level

Number of rows: 289

<input type="checkbox"/>	correction label	<i>approximate level</i>	Total
<input checked="" type="checkbox"/>	Σ	Σ	6,209.3 (918)
<input type="checkbox"/>	C,_L-W L-W L-W:2	Nybörjare	6.8 (1)
<input type="checkbox"/>	C,_L-W __UNDEF__ __UNDEF__	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-Der,_O,_O-Comp L-Der:2,_O:2,_O-Comp:2 L-Der:3,_O:...	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-Ref,_S-Adv __UNDEF__ __UNDEF__	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W C S-Clause	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W L-W:2 L-W:3 L-W:4	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W L-W L-W:2	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W M-Gend,_O O	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W M-Gend __UNDEF__	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W M-Num __UNDEF__	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W,_M-Verb S-Clause M-Def	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W O L-Der,_O,_S-Type	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W,_S-Type __UNDEF__ L-W,_S-Type __UNDEF__	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W,_S-Type __UNDEF__ M-Def	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W __UNDEF__ L-Der,_S-Type	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W __UNDEF__ L-W,_M-Num	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W __UNDEF__ M-Num	Nybörjare	6.8 (1)
<input type="checkbox"/>	L-W __UNDEF__ __UNDEF__ __UNDEF__	Nybörjare	6.8 (1)

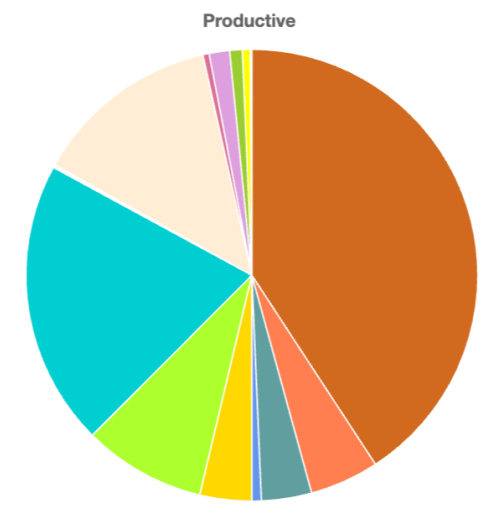
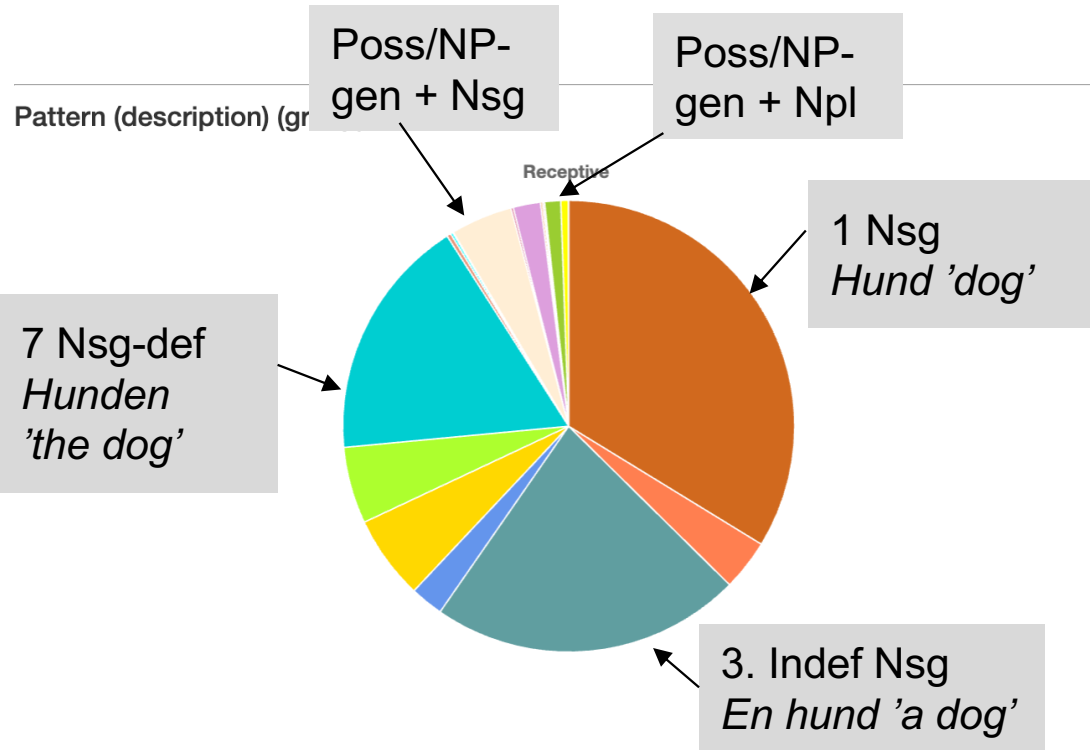
Nybörjare = beginner

SweLL gold original – correction labels & level

<input type="checkbox"/>	correction label	<i>▲ approximate level</i>	Total
<input checked="" type="checkbox"/>	__UNDEF__ __UNDEF__ __UNDEF__	Avancerad	 1,826.3 (270)
<input type="checkbox"/>	__UNDEF__ __UNDEF__ M-Num	Avancerad	 155.6 (23)
<input type="checkbox"/>	__UNDEF__ M-Gend __UNDEF__	Avancerad	 108.2 (16)
<input type="checkbox"/>	L-W __UNDEF__ __UNDEF__	Avancerad	 94.7 (14)
<input type="checkbox"/>	__UNDEF__ __UNDEF__ M-Def	Avancerad	 74.4 (11)
<input type="checkbox"/>	__UNDEF__ __UNDEF__ __UNDEF__ __UNDEF__	Avancerad	 47.3 (7)
<input type="checkbox"/>	__UNDEF__ __UNDEF__ __UNDEF__ __UNDEF__ UN...	Avancerad	 40.6 (6)
<input type="checkbox"/>	__UNDEF__ O-Comp O-Comp:2	Avancerad	 40.6 (6)
<input type="checkbox"/>	__UNDEF__ M-Num __UNDEF__	Avancerad	 40.6 (6)
<input type="checkbox"/>	__UNDEF__ M-Def __UNDEF__	Avancerad	 40.6 (6)
<input type="checkbox"/>	M-Verb __UNDEF__ __UNDEF__	Avancerad	 40.6 (6)
<input type="checkbox"/>	__UNDEF__ __UNDEF__ O-Comp	Avancerad	 27.1 (4)
<input type="checkbox"/>	__UNDEF__ __UNDEF__ O	Avancerad	 27.1 (4)
<input type="checkbox"/>	__UNDEF__ __UNDEF__ L-Der	Avancerad	 27.1 (4)
<input type="checkbox"/>	__UNDEF__ L-W __UNDEF__	Avancerad	 27.1 (4)
<input type="checkbox"/>	L-W __UNDEF__ M-Num	Avancerad	 27.1 (4)

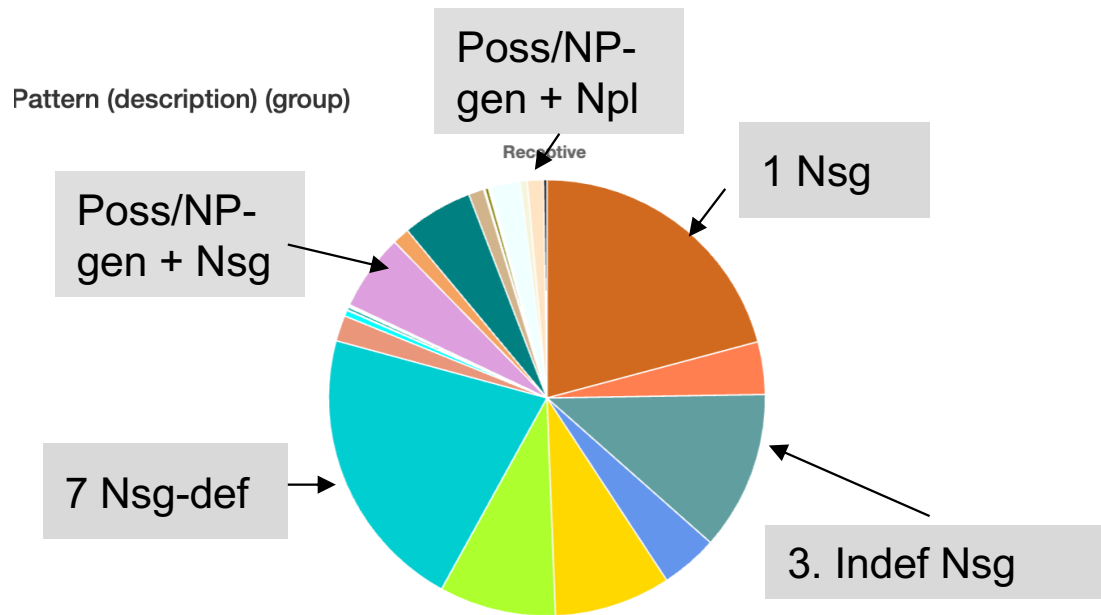
Avancerad = advanced

NB.
More poss./NP-gen. in the
productive data



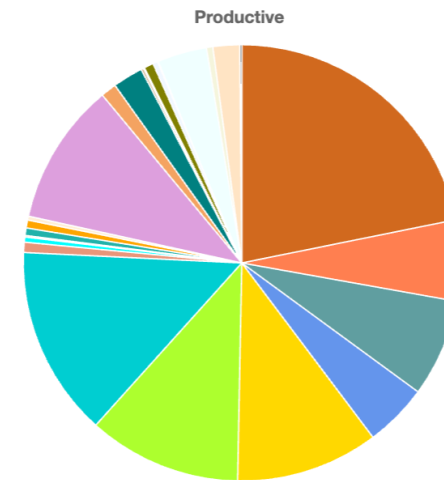
NP patterns

A1, varying position



Noun patterns

NB.
More poss./NP-gen. in the productive data



B1, varying position

Future work: Follow up questions

- Are NPs used in a normlike manner?
- How are learners learning NP patterns & rules?
- Are there enough complex NPs in the input?
 - Complex NPs often require agreement – number, gender, etc.
 - Agreement is important in assessment.
 - Some patterns rare and hard to learn?
- **Why are more NPs with a possessive modifier used in the productive data? (← A1 can-do's, topic,...)**

Noun patterns: Definiteness

Definiteness in Swedish – a challenge for learners

- Swedish has different means of expressing definiteness:
 1. **Definite suffix (main marker of definiteness):**
boken 'the book'

Definiteness in Swedish – a challenge for learners

- Swedish has different means of expressing definiteness:
 1. **Definite suffix (main marker of definiteness):**
boken 'the book'
 2. **A) Definite suffix and definite article (DOUBLE DEFINITENESS)** – compulsory with most preposed attributive modifiers:
den intressanta boken 'the interesting book'
 2. **B) Demonstrative pronoun *den här* (DOUBLE DEFINITENESS):**
den här boken 'this book'

Definiteness in Swedish – a challenge for learners

- Swedish has different means of expressing definiteness:
 1. **Definite suffix (main marker of definiteness):**
boken 'the book'
 2. **A) Definite suffix and definite article (DOUBLE DEFINITENESS):**
den intressanta boken 'the interesting book'
 2. **B) Demonstrative pronoun *den här* (DOUBLE DEFINITENESS):**
den här boken 'this book'
 3. **Possessive pronouns, genitives and demonstrative pronoun *denna* – without the definite suffix:**
min bok 'my book'
Thereses bok 'Therese's book'
denna bok 'this book'

- **Noun patterns (28+)**
(based on Agebjörn (unpubl.) and *Svenska Akademiens Grammatik* (SAG, Swedish Academy Grammar))
- All patterns divided according to **gender**: 28 x 2 = 56
- Some according to **position** – initial or any position → 56+
- All patterns tagged with:
 - pattern structure
 - clause position
 - definiteness
 - gender
 - number
 - attribute
 - article
 - other definite attribute
 → **filters**

	Singular					Plural					
	#	Art./pron./gen.	Adj. attr.	N	Def. suff.	#	Art./pron./gen.	Adj. attr.	N	Num. suff.	Def. suff.
IN-DEF.	3	en/ett		N	-	5	-		N	pl	-
	4	en/ett	A	N	-	6	-	A	N	pl	-
	1	-		N	-	5	-		N	pl	-
	2	-	A	N	-	6	-	A	N	pl	-
DEF.	7	-		N	-en/-et	17	-		N	pl	-na
	8	den/det	A	N	-en/-et	18	de	A	N	pl	-na
	11	den/det		N	-	21	de		N	pl	-
	12	den/det	A	N	-	22	de	A	N	pl	-
	9	den här/ den där		N	-en/-et	19	de här/ de där		N	pl	-na
	10	den här/ den där	A	N	-en/-et	20	de här/ de där	A	N	pl	-na
	13	denna		N	-	23	dessa		N	pl	-
	14	denna	A	N	-	24	dessa	A	N	pl	-
	15	min/Kalles		N	-	25	mina/Kalles		N	pl	-
	16	min/Kalles	A	N	-	26	mina/Kalles	A	N	pl	-
	27	-	e.g. förra, andra	N	-en/-et	28	-	e.g. förra, andra	N	pl.	-na

Nouns patterns – choose by pattern or filter



Noun patterns

Pattern (description) ▼ Pattern (structure) ▼ Clause position ▼ Definiteness ▼

Gender ▼ Number ▼ Attribute ▼ Article ▼ Other definite attributes ▼

A1 A2 B1 B2 C1 C2

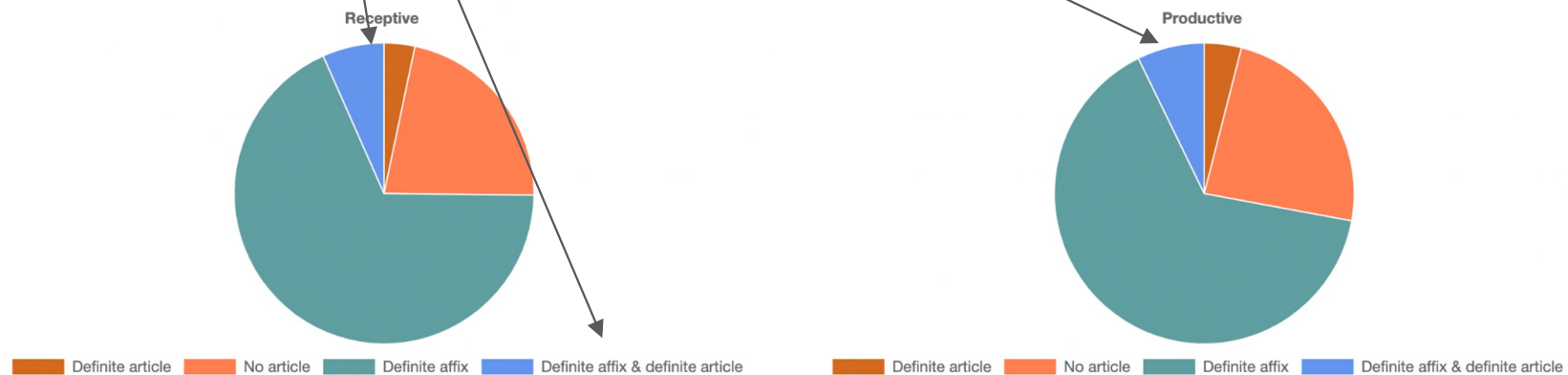
Tables Graphs Statistics Download Clear all

Extend columns in the table Show only first occurrence [Tables - description](#)  [Filters - description](#) 

Definite NPs only – 'articles' at B2/C1/C2





- Definiteness usually = definite affix (Sw. *slutartikel*)
- **Double definiteness** – only slightly more common than **only definite article**
den / det / de
- **No article forms** – include possessive pronouns, genitives, some demonstrative *denna* which take bare nouns.





Article

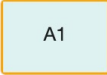
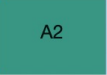
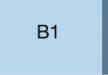
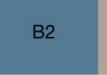
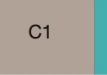





Demonstratives, possessives, genitives with adjectival attributive modifiers

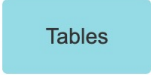

Noun patterns

Pattern (description)  Pattern (structure)  Clause position Varying  Definiteness 

Gender  Number  Attribute Preposed attribute  Article 


     

Show only first occurrence [Graphs - description](#)  [Filters - description](#) 

Tables  **Graphs** 

[Other definite attributes](#)

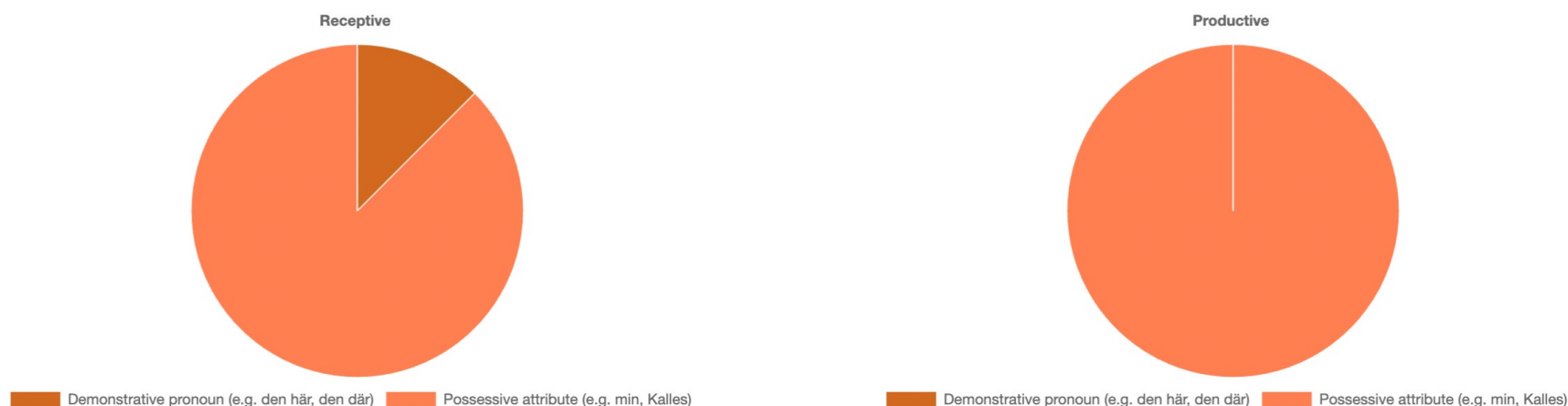
- Demonstrative pronoun (e.g. den här, den där)
- Demonstrative pronoun (e.g. denna)
- Possessive attribute (e.g. min, Kalles)
- No definite attribute



POSS / GEN / DEM.PRON. + ADJ.

A1 – no demonstratives in learner data

Other definite attributes



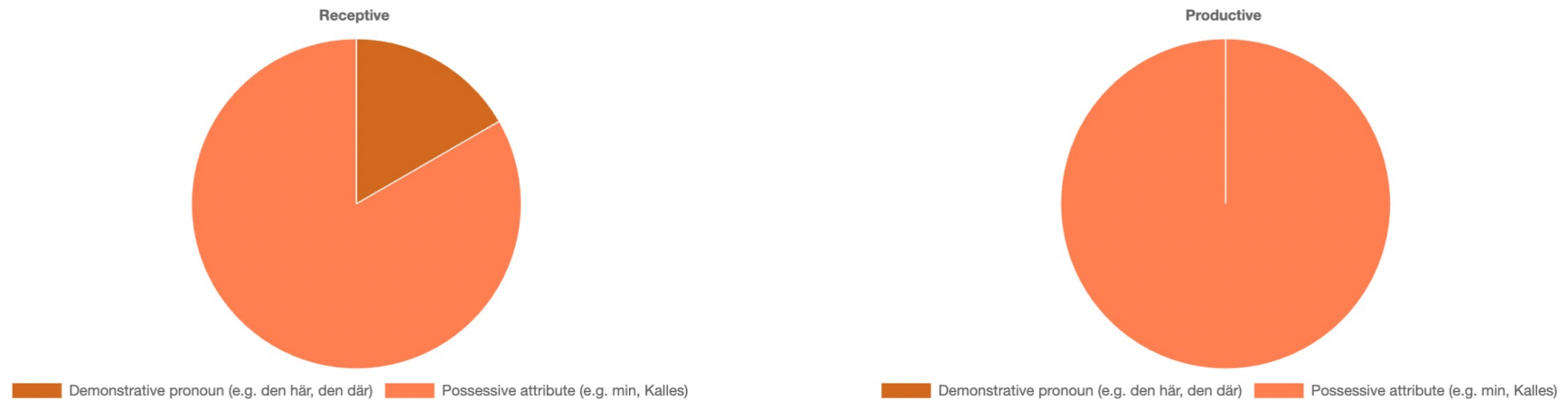
A1 – no demonstratives in learner data – not surprising

A1: “Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has.” (COE 2001, p. 24)

→ focus on personal → possessive pronouns

A2 – still no demonstratives in learner data

Other definite attributes

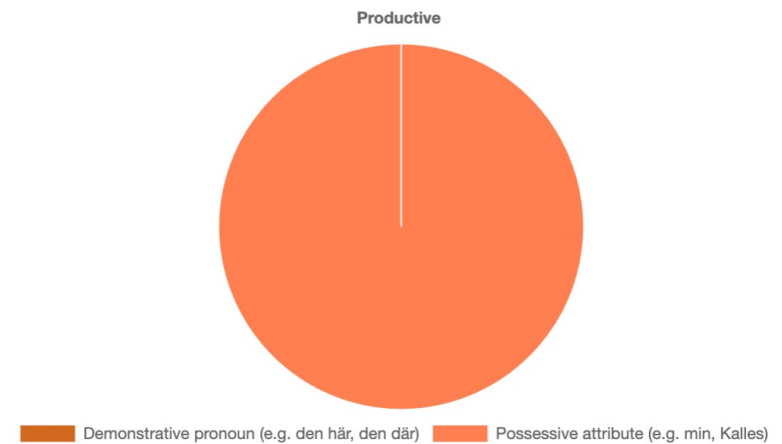
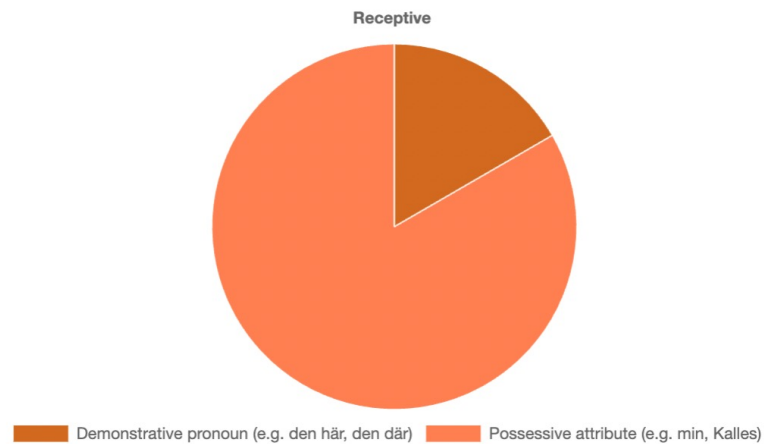


A2: “Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.” (COE, 2001, p. 24)

→ Writing tasks → lack of demonstratives

A2 – still no demonstratives in learner data

Other definite attributes



Are they really not producing any demonstratives

OR are they combining them with the wrong morphological form?

Non-norm usage?

Noun patterns [🔗](#)

Pattern (description) Pattern (structure) Clause position Definiteness

Gender Number Attribute Article Other definite attributes

Preposed attribute Demonstrative pronoun...

A1 A2 B1 B2 C1 C2

Tables Graphs Statistics Download Clear all

Extend columns in the table Show only first occurrence [Tables - description](#) [Filters - description](#)

Total rows : 3

CEFR level <input type="text"/>	Pattern (description) <input type="text"/>	Pattern (structure) <input type="text"/>	Receptive <input type="text"/>	Productive <input type="text"/>
A2	20b. de härl dår A Npl-def (<i>de här gröna knapparna</i>)	Complex noun phrase	0.02 (3)	0.00 (0)
A2	10c. det härl dår A Nsg-def (<i>Det här speciella förortsspråket</i>)	Complex noun phrase	0.01 (1)	0.00 (0)
A2	10b. den härl dår A Nsg-def (<i>Den där nya deckaren</i>)	Complex noun phrase	0.03 (4)	0.00 (0)

CLICK
Access corpus data + other corpora

Change the query → non-norm → still no hits



3 of 275 corpora selected — 160.06K of 15.19G tokens

SPRÅKBANKENTEXT

Simple Extended **Advanced** Compare

Active CQP query in simple search:

Active CQP query in extended:

[]

Custom CQP query:

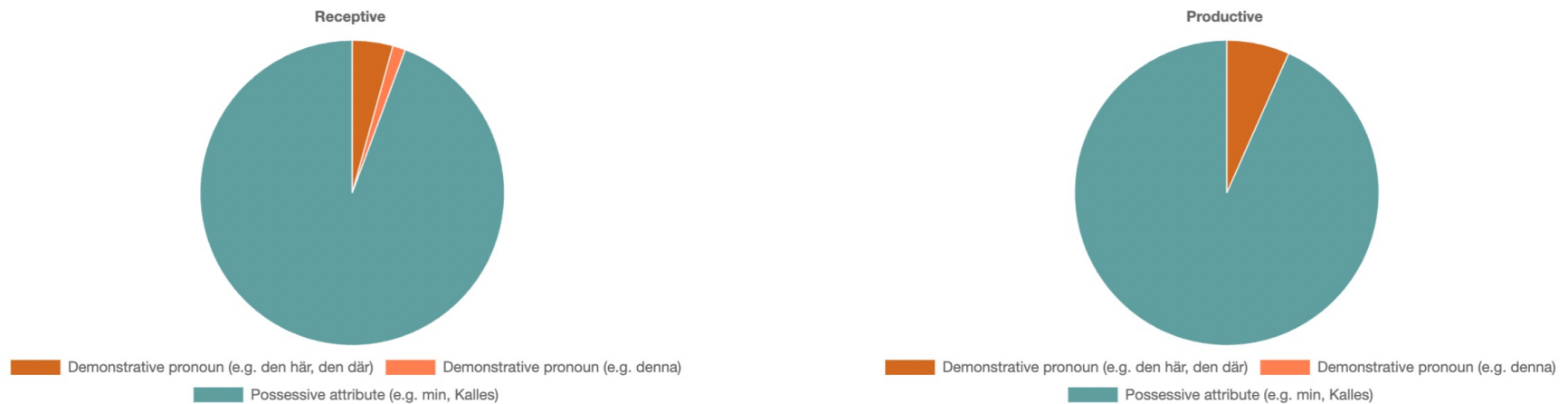
[Get the CQP tutorial](#)

```
JJ | pos = PC | pos = RV | lex contains  
"andra\.\.av\1" | lex contains "annan\.\.pn\1"] [word = ""{0,1} [(msd = "NN\.\.UTR\.\.SIN\.\.DEF\.\.NOM") & pos = "NN" & lex not
```


B1

B1: “Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.” (COE, 2001, p. 24)

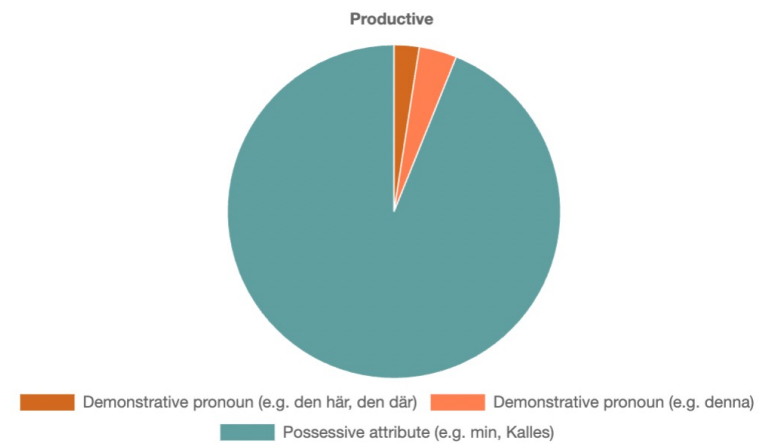
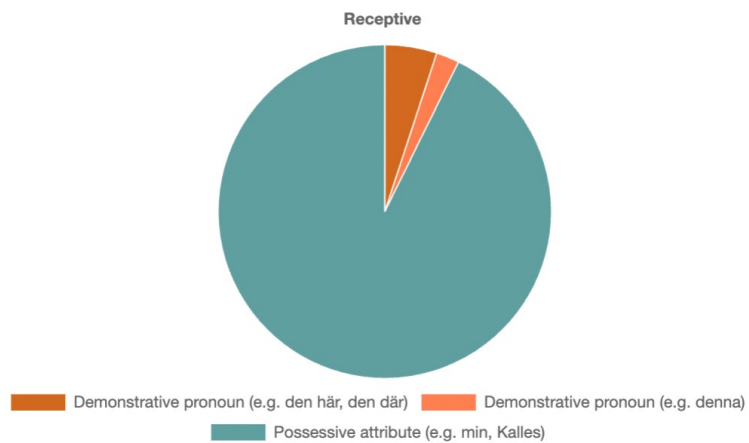
Other definite attributes



- Some demonstrative pronoun *den här* 'this' in the learner data
- Demonstrative *denna* 'this' new at this level in course books – more formal than *den här*

B2

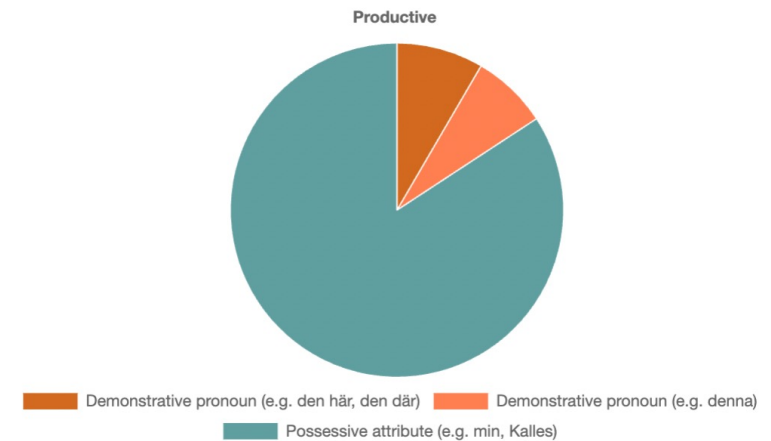
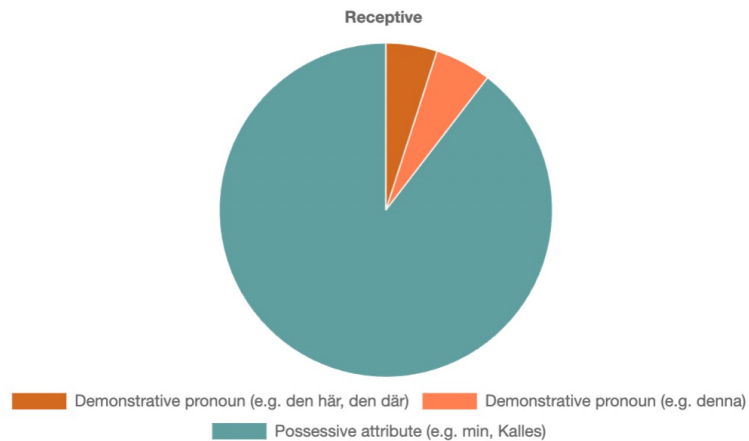
Other definite attributes



- *Denna* 'this' new in learner data. – *Denna* is more common in formal writing, and in certain regions.
- More common than *den här* 'this' in learner data --> Why?
Overuse of *denna*? Due to non-normative usage of *den här*?

C1

Other definite attributes



- More demonstratives in receptive and productive.
- *Denna* 'this' & *den här* 'this' equal proportions

Previous research on definiteness

NOTE:

Our patterns are norm-based but might be used in non-norm contexts

Nyqvist 2018:19 – grammaticality judgement test, double definiteness, immersion (16 years old) vs non-immersion students (15 years old)

Acquisition order (75 % accuracy):

(1) **Double definiteness:** common singular > neuter singular > plural

(2) **Definiteness with PRG attributes:** *den här* > genitive attribute with adj. > possessive attribute with adj. > *denna*

- *Den här* 'this' occurs with the definite suffix – the most common way of marking definiteness.
- *Denna* 'this' is quite rare in input *and* occurs with the more unusual bare noun.

Comparison with reference corpora
→ useful for e.g. DDL

Comparison with L1 reference corpora easily done

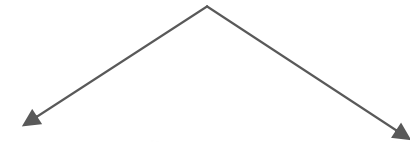
- Reference corpora **not visible inside the profile**, but...
- All of our patterns can be easily searched also in reference corpora available in Korp, Språkbanken Text, Gothenburg.
- No overview (future work but requires downloading queries)
- Let's look at one pattern:
pattern 1b – hard for learners to know when to use the 'naked' indefinite form

*Många köpte **hund** under covid.*
many buy-PT **dog** during covid
“Many bought **a dog** during covid”

Frequency – link to search → reference corpora

Click on the frequency → Hits in Korp → can change the corpus to an L1 reference corpus

*Många köpte **hund** under covid.*
many buy-PT **dog** during covid
“Many bought **a dog** during covid”



CEFR level ↕	Pattern (description) ↕	Pattern (structure) ↕	Receptive ↕ 📄	Productive ↕ 📄
A1	1b. Nsg (Jag kan svenska)	Simple noun phrase	1.00 (76)	1.42 (9)
A2	1b. Nsg (Jag kan svenska)	Simple noun phrase	1.04 (145)	1.41 (37)
B1	1b. Nsg (Jag kan svenska)	Simple noun phrase	0.74 (137)	1.57 (27)
B2	1b. Nsg (Jag kan svenska)	Simple noun phrase	0.83 (139)	2.75 (74)
C1	1b. Nsg (Jag kan svenska)	Simple noun phrase	0.89 (148)	3.26 (94)
C2	1b. Nsg (Jag kan svenska)	Simple noun phrase	NaN (0)	2.86 (5)

Custom CQP query:

[Get the CQP tutorial](#)

```
[(msd = "NN\.UTR\,SIN\,IND\,NOM") & pos = "NN" & lex not contains "januari\.\.nn\,1" & lex not contains "februari\.\.nn\,1" & lex not contains "mars\.\.nn\,1" & lex not contains "april\.\.nn\,1" & lex not contains "maj\.\.nn\,1" & lex not contains "juni\.\.nn\,1" & lex not contains "juli\.\.nn\,1" & lex not contains "augusti\.\.nn\,1" & lex not contains
```

Search

KWIC: hits per page: 25 sort within corpora: not sorted Statistics: compile based on: word Show statistics Show word picture

[KWIC](#) [Statistik](#) [Ordbild](#)

Results: 148

« 1 2 3 4 5 6 » Go to page of 6 [Show context](#)

COCTAILL LESSON TEXT

Straat låter	det vara varmt där, han låter kvinnorna vara vackra, de fattiga fattiga och de rika rika.
Uppfostran blev	snäsor och luggar, » Gud som haver » och vara lydig.
Mamma kommer	ut.
Lunch i	morgon?
Finska talas	av ca 250 000 personer i Sverige.
Mamma tog	några bestämda steg mot vardagsrummet.
Skilsmässa är	en vanlig orsak men många föräldrar är redan från början ensamboende.

Change to a reference corpus

- e.g. newspapers GP 2013 (Göteborgsposten)
- Remove the level from the search

KORP v9  2 of 275 corpora selected — 652.00K of 15.19G tokens

SPRÅKBANKENTEXT  SWE-CLARIN [(msd = "NN\,UTR\,SIN\,IND\,NOM") & pos = "NN" & lex not contains "januari\.\.nn\,1" & lex not contains "februari\.\.nn\,1" & lex not contains "mars\.\.nn\,1" & lex not contains "april\.\.nn\,1" & lex not contains "maj\.\.nn\,1" & lex not contains "juni\.\.nn\,1" & lex not contains "juli\.\.nn\,1" & lex not contains "augusti\.\.nn\,1" & lex not contains ...]

Simple Extended **Advanced** Compare

Active CQP query in simple search:

Active CQP query in extended:
[]

Custom CQP query:
[Get the CQP tutorial](#)
[(msd = "NN\,UTR\,SIN\,IND\,NOM") & pos = "NN" & lex not contains "januari\.\.nn\,1" & lex not contains "februari\.\.nn\,1" & lex not contains "mars\.\.nn\,1" & lex not contains "april\.\.nn\,1" & lex not contains "maj\.\.nn\,1" & lex not contains "juni\.\.nn\,1" & lex not contains "juli\.\.nn\,1" & lex not contains "augusti\.\.nn\,1" & lex not contains ...]

Remove the level

Newspaper: *Trial started...; Woman raped...; Traffic planning must...*

KWIC

Statistics

Word picture

Results: 21,142

« < 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 ... > »

Go to page of 846 [Show context](#)

GP 2013

Rättegång inleds	i New Orleans i ett civilmål mot oljebolaget BP anklagat för att inte ha tagit katastrofen vid den läckande oljekällan i Mexikanska golfen 2010 på allvar.
Kvinna våldtagen	
Trafikplanering måste	vara framåtriktad.
Folkbildning i	ett väldigt fint format.
10-åring överfölls	av man i bil
Hörna för	Gais, Muhrén som skytt – och Niklas Hellqvist i Sirius som fullföljde efter att skottet avlossats.
Stress och	krav på snabbare flöden genom förlossningsavdelningen skulle med andra ord kunna leda till att de kloka riktlinjer som arbetats fram inom projektet ”S
Raggare och	Serotonin är två låtar av krautpionjärerna Can, med svenska texter, ett sound som både skaver och har rymd.
Rektor har	sagt att hon ska se vad hon kan göra, säger Josefin Krantz.
Confitering är	ett gammalt sätt att konservera och laga till anklåren.
Polis kontrollerar	en bil vid Rundradiogatan.
Minister misstänkt	för plagiat avgår
ordförande i	SKL:s utbildningsberedning

Comparing levels, comparing corpora...

How can we compare these?

- Currently relative frequency in ***Swedish L2 Grammatical Profile*** is **occurrences / 100 sentences**
 - Total tokens → not the best for grammatical constructions, important to relate to sentences, clauses or phrases.
 - Verb phrases first → per 100 sentences
- Future research: **Which is the best relative measure?**
 - Nouns – relative in comparison to the total number of nouns?
 - Verbs – relative in comparison to the total number of verbs?

Preliminary comparison (newspaper & learner essays)

	GP 2013 (newspaper)	A1 prod.	A2 prod.	B1 prod.	B2 prod.	C1 prod.
1b Nsg (per 1 mil. tok.)	1253.1 (21142)	1740.1	1355.1	1086.4	1619.8	1806.0
1b Nsg (per 100 sent.)	1.69	1.42 (9)	1.41 (37)	1.57 (27)	2.75 (74)	3.26 (94)
3b Indef Nsg (per 1 mil. tok.)	8091.2 (136 515)	5800.5	6519.2	7162.1	6523.1	6897.5
3b Indef Nsg (per 100 sent.)	10.91	4.74 (30)	6.80 (178)	10.35 (178)	11.07 (298)	12.44 (359)
7c Nsg-DEF (per 1 mil. tok.)	4375.9 (73830)	2706.9	2014.4	1609.5	1072.6	1940.5
7c Nsg-DEF (per 100 sent.)	5.9	2.21 (14)	2.10 (55)	2.33 (40)	1.82 (49)	3.50 (101)

Possible **underuse** of all patterns, except 1b & 3b from B2 (per 100 sentences).
 1b, 3b possible signs of **overuse** from B2. → needs further study.

Future research

- Best way of normalising the frequencies...
- Development of non-normative patterns per level
→ patterns in the non-normative usage
- Could more norm-like usage be stimulated by increased usage in e.g. coursebooks?
- Data Drive Learning

Swedish L2 profile – providing possibilities...

Teaching

Learning

Assessment

Research

- Open access to empirical data (learner data behind access license)
- Open access to reference corpora, possible to search in exactly the same way.
- Further analysis needed.
- ICALL exercises based on the material should be explored.

Underway:

- L2 Swedish development prepositions, passives, past tenses, definite NPs etc
- Prefixes
- Morphological complexity



References

- Bybee, J. 2008. Usage-based grammar and second language acquisition. In: P. Robinson & N.C. Ellis (red.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (pp. 216–236). New York: Routledge..
- Caines, A., & P. Buttery. 2017. The effect of task and topic on opportunity of use in learner corpora. *Learner corpus research: New perspectives and applications*, 5-27.
- Council of Europe (COE). 2001. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Press Syndicate of the University of Cambridge.
- Nyqvist, E-L. 2018. Mastering complex Swedish NPs: A comparison of non-immersion pupils and immersion L1 Finnish pupils. *Journal of the European Second Language Association*, 2 (1), 14–23.
- Teleman, U., Hellberg, S., & Andersson, E. (1999). *Svenska akademiens grammatik (SAG)*. Svenska akademien.
- Volodina, E., Pilán, I., Rødven Eide, S., Heidarsson, H. 2014. You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a second language. I: Fourth workshop on NLP for Computer Assisted Language Learning (NLP4CALL), ACL anthology, Linköping University Press.
- Volodina, E., Pilán, I., Enström, I., Llozhi, L., Lundkvist, P., Sundberg, G., & Sandell, M. 2016. SweLL on the rise: Swedish Learner Language corpus for European Reference Level studies. *Proceedings of LREC 2016, Slovenia*
- Volodina, E, Granstedt, L., Matsson, A., Megyesi, B., Pilán, I. , Prentice, J., Rosén, D., Rudebeck, L., Schenström, C-J., Sundberg, G. & Wirén, M. (2019). The SweLL Language Learner Corpus: From Design to Annotation. *Northern European Journal of Language Technology, Special Issue*.
- Volodina, E., Alfter, D., Lindström Tiedemann, T., Lauriala, M. S., & Piipponen, D. H. (2022). Reliability of automatic linguistic annotation: native vs non-native texts. In *Selected papers from the CLARINAnnual Conference 2021*. Linköping University Electronic Press (LiU E-Press).

Swedish L2 Profile



Exercise Generator Tools Experimental prototypes Swedish Profile

LARK   <https://spraakbanken.gu.se/larkalabb/svlp>

Language Acquisition Reusing **Korp**

Lexical profile Grammatical profile Morphological profile

<https://spraakbanken.gu.se/larkalabb/svlp> (demo)

Workshop profiling L2 lexis and grammar, (20 April 2023)

Therese Lindström Tiedemann*

therese.lindstromtiedemann@helsinki.fi

Based on work done with: Elena Volodina¹, Yousuf Ali Mohammed*¹, Anders Agebjörn, Maisa Lauriala*



UNIVERSITY OF GOTHENBURG

SPRÅKBANKENTEXT



UNIVERSITY OF HELSINKI