Swedish L2 profile -
a tool for exploring L2 data

Elena Volodina (presenter), University of Gothenburg, Sweden
Yousuf Ali Mohammed, University of Gothenburg, Sweden
Therese Lindström Tiedemann, University of Helsinki, Finland

L2 profiles for Swedish

L2 profiles for Swedish

L2P - Documentation

L2P - MWE experiment

L2P - experiment with word difficulty

Full name: Development of lexical and grammatical competences in immigrant Swedish, RJ, 2018-2020

Background

Sweden has a growing number of immigrants, the need for courses and coursebooks in Swedish as a second language (L2) is increasing, as is the demand for standardized tests and qualifications. This project intends to study the development of lexical and grammatical competences in L2 learners of Swedish.

General description

We intend to perform the study through two corpora: coursebook texts and learner essays, both marked up for proficiency levels according to the Common European Framework of References (CEFR). The corpora will be processed by computational methods, after which the results will be analysed by linguists, lexicographers, grammarians, teachers and language assessors - both linguistically, and based on theory of teaching, to find ways of identifying minimal or central (need-to-know) vocabulary and grammar scopes, as well as

Project duration
2018-01-01 – 2021-12-31

Project members

Elena Volodina
(Projektledare)
elena.volodina@svenska.gu.se

Therese Lindström Tiedemann (Researcher)
University of Helsinki

Yousuf Ali Mohammed
(Systems developer)
yousuf.ali.mohammed@svenska.gu.se

David Alfter (Systems developer (part-time) & PhD student (part-time))
david.alfter@svenska.gu.se

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

What is Swedish L2 profile?

Inspired by other profiles...

English Vocabulary Profile Online - British English

Search...

Topic

body and health

Hide culturally sensitive

Results 1 - 20 of 356

Base Word	Guideword
medicine	SUBSTANCE
heart	ORGAN

English Grammar Profile Online

Search...

Clear Results

Display #

Details

Details

Details

PEARSON ENGLISH › Global Scale of English › GSE Teacher Toolkit

GSE Teacher Toolkit

Let the Global Scale of English (GSE) save you time and improve the quality and relevance of your English classes. The GSE Teacher Toolkit makes it quick and easy to find learning objectives, grammar, vocabulary and coursebook recommendations all mapped to the GSE and CEFR.

Play Overview

Choose category ❓

Learning Objectives	Grammar	**Vocabulary**	Text Analyzer

Who are you teaching?

Choose Learner ▾

Choose a range on the GSE / CEFR ❓

	< A1	A1	A2	A2+	B1	B1+	B2	B2+	C1	C2
10										**90**
22	30	36	43	51	59	67	76	85		

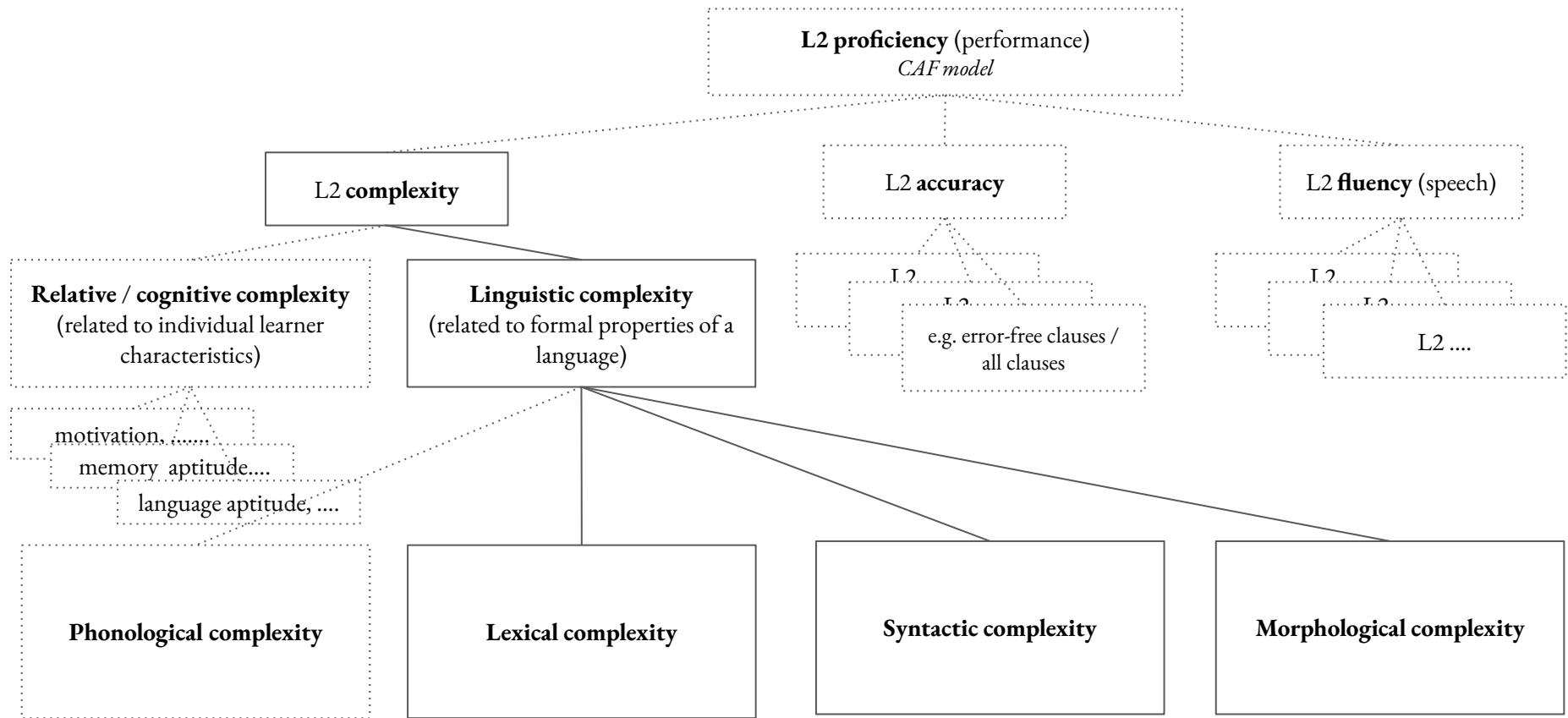🔖 Choose Topic ❓   📄 Grammatical Category ❓   🔍 Filter search results with an exact word

Show results

Why?

- Need to visualize and make sense of data
- Generate hypotheses
- Link numbers to actual data samples
- …

L2 proficiency (performance)
CAF model

L2 complexity

L2 accuracy

L2 fluency (speech)

Relative / cognitive complexity
(related to individual learner characteristics)

Linguistic complexity
(related to formal properties of a language)

L2 ....

L2 ....

e.g. error-free clauses / all clauses

L2 ....

L2 ....

L2 ....

motivation, ...

memory aptitude....

language aptitude, ...

Phonological complexity

Lexical complexity

Syntactic complexity

Morphological complexity

Swedish L2 Profile

Language Acquisition Reusing Korp

Lexical profile	Grammatical profile	Morphological profile
Adjectival declension	Verb patterns	Word family
Adjectival & Adverbial struct	Noun patterns	Morpheme family
Multi Word Expressions		
Sen*Lex		

Note:
not inflectional morphology

Swedish L2 profile

(will be) freely available at (login 'demo'): https://spraakbanken.gu.se/larkalabb/svlp

live demo:

https://spraakbanken.gu.se/larkalabb/svlp

DEMO

- Lexical profile
  - Sen*Lex: conjunctions/subjunctions, gender
  - MWEs (1st occurrence vs usage, per CEFR); verbal types
- Grammar profile
  - Verbs: Preterite vs Past
  - Nouns: in general vs incl. attributes
- Morphological profile
  - Morpheme family: word formation mechanisms
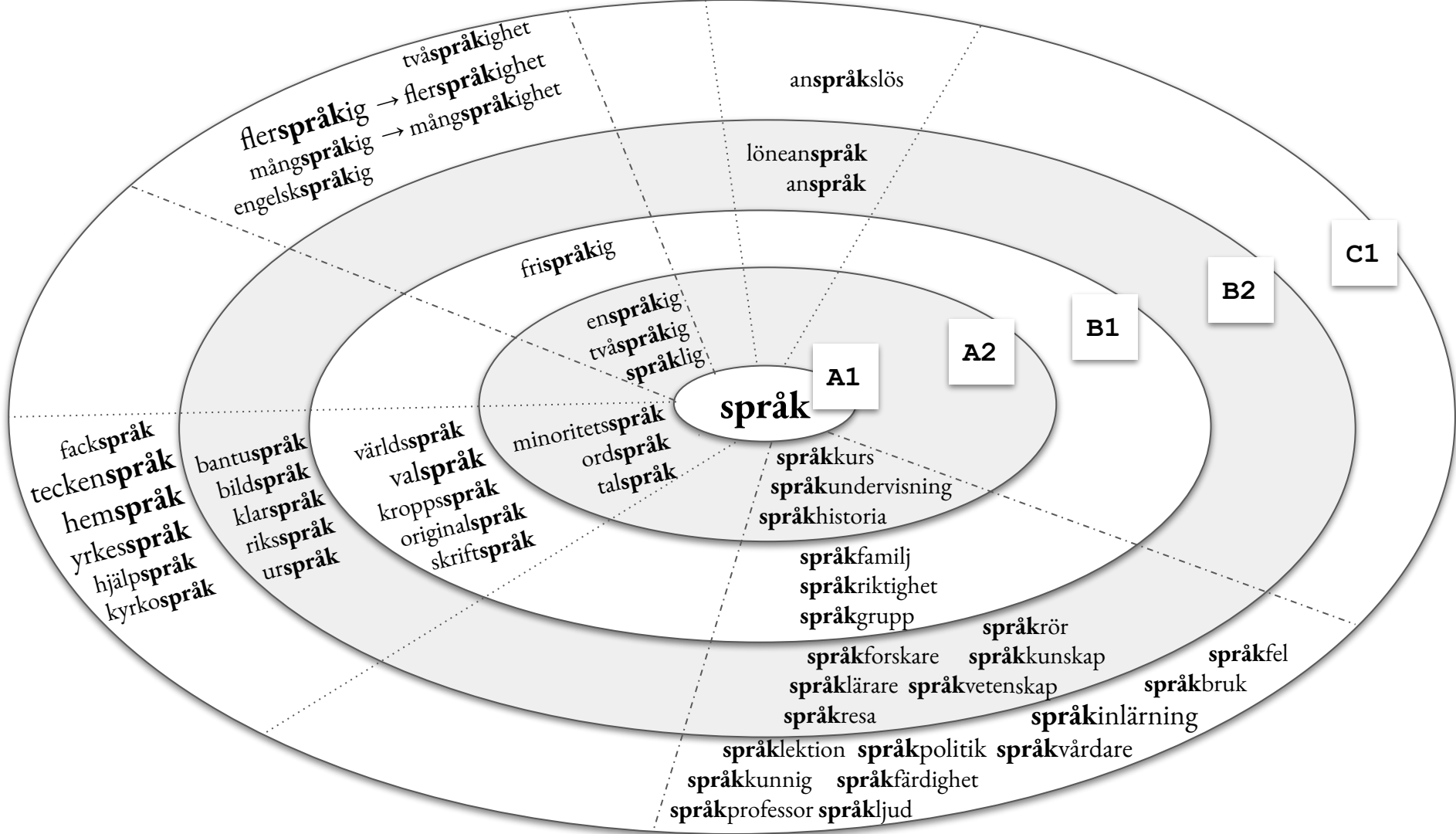 - Word family: dag, språk, lyx

A1 (17 members) →   A2 (36 members) →   B1 (28 members) →   B2 (13 members) →   C1 (7 members)

beginner   → elementary   → intermediate   → upper-intermediate   → advanced

A1   A2   B1   B2   C1

språk

 enspråkig
tvåspråkig
språklig

frispråkig

minoritetsspråk
ordspråk
talspråk

anspråkslös

löneanspråk
anspråk

tvåspråkighet
flerspråkig → flerspråkighet
mångspråkig → mångspråkighet
engelskspråkig

världsspråk
valspråk
kroppsspråk
originalspråk
skriftspråk

bantuspråk
bildspråk
klarspråk
riksspråk
urspråk

fackspråk
teckenspråk
hemspråk
yrkesspråk
hjälpspråk
kyrkospråk

språkkurs
språkundervisning
språkhistoria

språkfamilj
språkriktighet
språkgrupp

språkrör
språkforskare   språkkunskap
språklärare   språkvetenskap   språkfel
språkresa   språkbruk
språkinlärning
språklektion   språkpolitik   språkvårdare
språkkunnig   språkfärdighet
språkprofessor   språkljud

lyx ('luxury')

```
hus ('house') (A1) →
varuhus ('store, house of products') (A1) →
vara ('product, item') (A2) →
lyxvara ('item of luxury') (B1) →
lyx ('luxury') (B2)
```
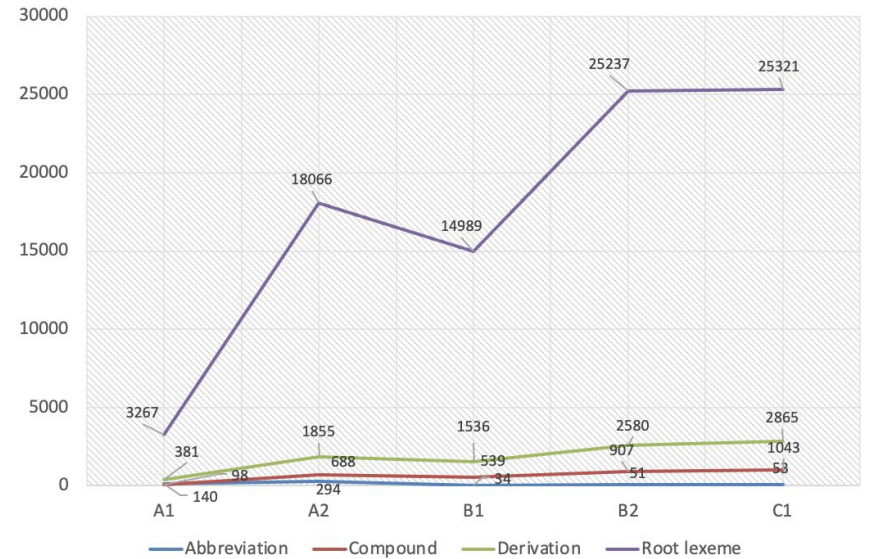
Root morpheme ↓↑	CEFR level ↓↑	Lemgram ↓↑
lyx	B1	lyxvara..nn.1
lyx	B2	lyx..nn.1
lyx	B2	lyxa..vb.1
lyx	B2	lyxig..av.1
lyx	C1	lyxliv..nn.1

Relate statistics and levels of language development

Word formation mechanisms in COCTAILL data per level

- A1: 16488, 2671, 1179, 62
- A2: 41744, 5032, 1856, 104
- B1: 71287, 8405, 2986, 168
- B2: 75231, 7497, 2728, 119
- C1: 71975, 7366, 2501, 232

Legend: Abbreviation, Compound, Derivation, Root lexeme
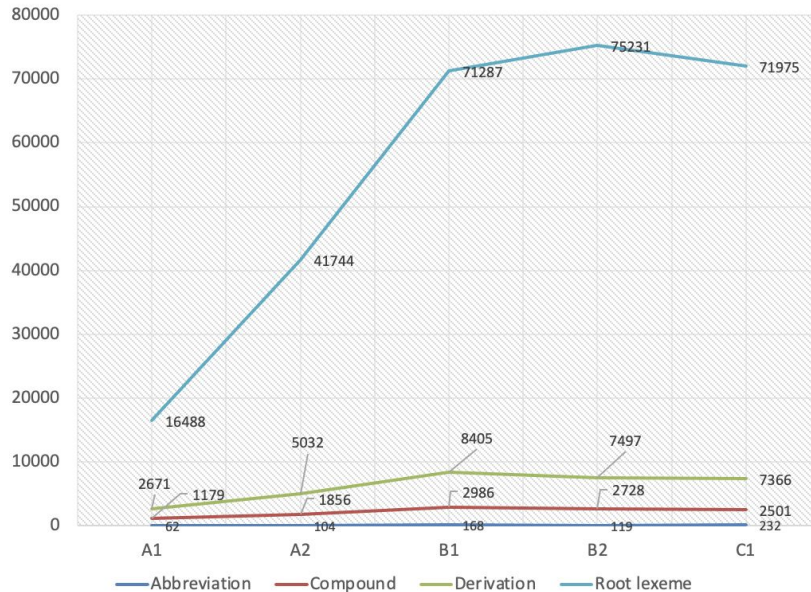
Word formation mechanisms in SweLL-pilot essay data per level

- A1: 3267, 381, 98, 140
- A2: 18066, 1855, 688, 294
- B1: 14989, 1536, 539, 34
- B2: 25237, 2580, 907, 51
- C1: 25321, 2865, 1043, 53

Legend: Abbreviation, Compound, Derivation, Root lexeme

Swedish L2 profile – providing possibilities…

Teaching | Learning | Assessment

- Open access to empirical data (learner data behind access license)
- Open access to reference corpora, possible to search in exactly the same way.

- ICALL exercises definite possibility, still to be explored.

Research

- L2 Swedish development, e.g. prepositions, passives, etc
- Word / morpheme families
- Morphological complexity

Thank you!
Comments? Questions?

(or mailto: elena dot volodina at svenska dot gu dot se)

| Lexical profile | Grammatical profile | Morphological profile |

Sen*Lex ⧉

| Search word | Word Class
Noun (NN) ▾ | Saldo Word Class ▾ | Noun declension ▾ | Gender ▾ |

| A1 | A2 | B1 | B2 | C1 | C2 | | Tables | Graphs | Statistics | Download | Clear all |

☐ **Show only first occurrence**   *Tables - description* ⧉   *Filters - description* ⧉

Total rows : 2221

CEFR level ⇅	Word ⇅	Lemgram ⇅	Sense ⇅	Word Class ⇅	Saldo Word Class ⇅	Receptive ⇅ ⧉	Productive ⇅ ⧉
A2	skola	skola..nn.1	skola..1	Noun (NN)	Noun (nn)	11.92 (82)	93.53 (214)
A2	kompis	kompis..nn.1	kompis..1	Noun (NN)	Noun (nn)	2.76 (19)	62.94 (144)
A2	dag	dag..nn.1	dag..1	Noun (NN)	Noun (nn)	23.54 (162)	52.45 (120)
A2	år	år..nn.1	år..1	Noun (NN)	Noun (nn)	37.35 (257)	40.21 (92)

17

Noun declension

Receptive

- 1. -or (blomma, blommor)
- 2. -ar (älv, älvar)
- 3. -er (dans, danser; (bok, böcker))
- 4. -r (sko, skor)
- 5. -n (bo, bon; nöje, nöjen)
- 6. - (bär, bär)
- Other (lexikon, lexika)

Productive

4. -r (sko, skor): 6

- 1. -or (blomma, blommor)
- 2. -ar (älv, älvar)
- 3. -er (dans, danser; (bok, böcker))
- 4. -r (sko, skor)
- 5. -n (bo, bon; nöje, nöjen)
- 6. - (bär, bär)
- Other (lexikon, lexika)
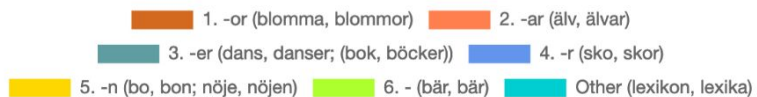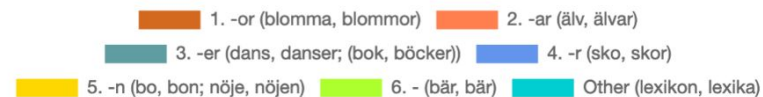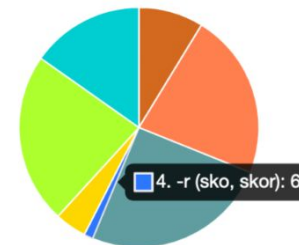
Noun declension	Receptive
1. -or (blomma, blommor)	211
2. -ar (älv, älvar)	475
3. -er (dans, danser; (bok, böcker))	572

Noun declension	Productive
1. -or (blomma, blommor)	43
2. -ar (älv, älvar)	110
3. -er (dans, danser; (bok, böcker))	124

Gender

Receptive

Legend: Common (en älv, en bil) | Common or neuter | Neuter (ett ägg, ett år) | No gender | Always plural - no gender
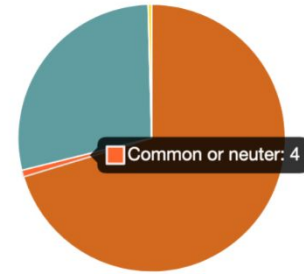
Productive

Common or neuter: 4
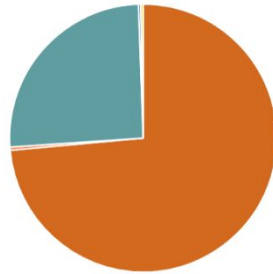
Legend: Common (en älv, en bil) | Common or neuter | Neuter (ett ägg, ett år) | No gender | Always plural - no gender

Gender	Receptive
Common (en älv, en bil)	1537
Common or neuter	8
Neuter (ett ägg, ett år)	533

Gender	Productive
Common (en älv, en bil)	346
Common or neuter	4
Neuter (ett ägg, ett år)	140

Total rows : 649

CEFR level	Word	Lemgram	Sense	Word Class	Saldo Word Class	Receptive	Productive
A1	skola	skola..nn.1	skola..1	Noun (NN)	Noun (nn)	11.20 (29)	249.02 (102)

Antal träffar: 102

SPIN (stödjer ej utökad kontext)

min **skolan** ligger på N-gata på N-gata.

I **skolan** är 300 elever.

NN arbetar i **skolan** .

Ja bra i **skolan** pratar svenska.

Min **skolan** heter NN.

Min **skola** ligger på N-gata på N-gata.

Det finns fyra våningar i **skolan** .

i **skolan** eleverna finns 16.

Jag slutar **skolan** kl. 2:00.

Jag kommer **skolan** sprbangn och but.

Jag tycker om bild och Lesnär på music och Simu **skola** .

20

Morpheme family

Search morpheme
ling

Search word
Search word

Morpheme Category
- ☐ Binding morpheme
- ☐ Derivational prefix
- ☑ Derivational suffix
- ☐ Inflectional ending
- ☐ Root morpheme
- ☐ Unclassified / unknown

Word Class ▼

Saldo Word Class ▼

Word formation ▼

A1	A2	B1	B2	C1	C2

Tables Graphs Statistics Download Clear all

☐ Add column for related roots   ☑ Show only first occurrence   *Tables - desc...*

Total rows : 12

Morpheme ⇅	CEFR level ⇅	Morpheme Category ⇅	Lemgram ⇅	Sense ⇅	Word Class ⇅	Saldo Word Class ⇅	Word formation ⇅	Receptive ⇅	Productive ⇅
ling	A1	Derivational suffix	älskling..nn.1	älskling..1	Noun (NN)	Noun (nn)	Derivation	1.16 (3)	0.00 (0)
ling	A2	Derivational suffix	Moderata_samlingspartiet..pmm.1	Moderata_samlingspartiet..1	Adjective (JJ)	Multiword proper noun (pmm)	Lexicalized form	0.29 (2)	0.00 (0)
ling	A2	Derivational suffix	böckling..nn.1	böckling..1	Noun (NN)	Noun (nn)	Derivation	0.29 (2)	0.00 (0)
ling	B1	Derivational suffix	änkling..nn.1				Noun (nn)		

Tables - des...

Lemgram ⇅

älskling..nn.1

lingon..nn.1

lingonsylt..nn.1

Moderata_sam

älskling

Definition :
- älskling (svenska.se)

Morphology :
- r: älsk s: ling

SAOL Analysis :
- älsk·ling

SO Analysis :
- älsk·ling·en

Antal träffar: 3

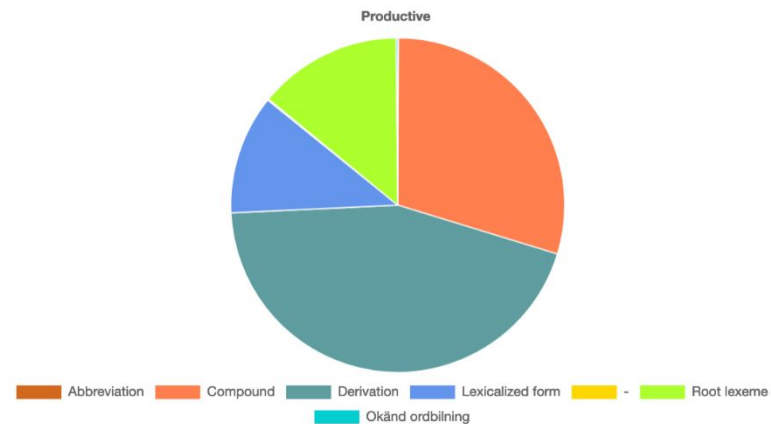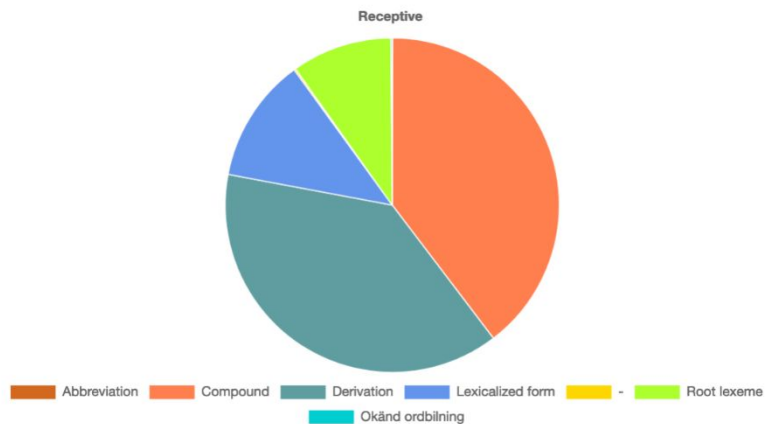« ‹ 1 › »   Gå till sida [   ] av 1   Visa kontext

COCTAILL

Mona: God natt, älskling, och sov gott!

Lasse: Hej, älskling, nu är jag hemma!

Lasse: God morgon, älskling!

« ‹ 1 › »   Gå till sida [   ] av 1

21

Word formation mechanisms

Receptive

Word formation	Receptive
Abbreviation	17
Compound	14344
Derivation	13871
Lexicalized form	4367

Productive

| Word formation	Productive
Abbreviation	4
Compound	1629
Derivation	2448
Lexicalized form	634

Statistics: type-token ratio

CEFR level	Receptive (Type)	Productive (Type)	Receptive (Token)	Productive (Token)	Receptive (TTR)	Productive (TTR)
A1	2166	464	25015	4083	0.09	0.11
A2	5134	714	16367	1866	0.31	0.38
B1	10327	1184	27224	2829	0.38	0.42
B2	10916	1310	22569	2748	0.48	0.48

Type-token ratio over all morphemes
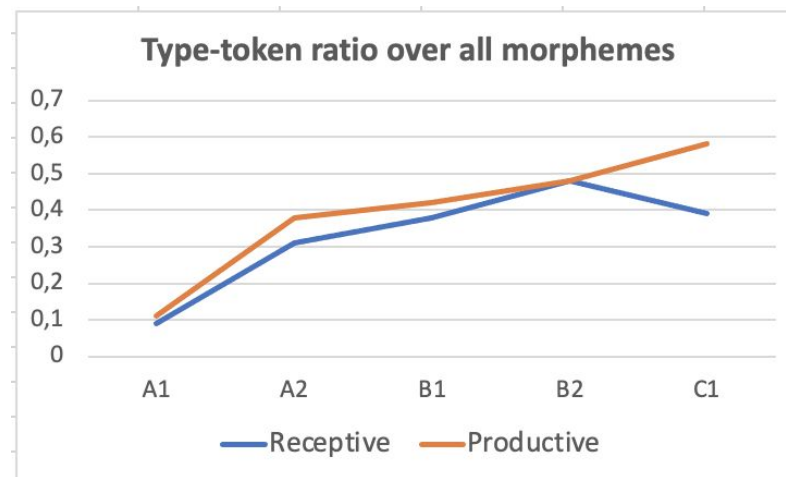
Morpheme categories
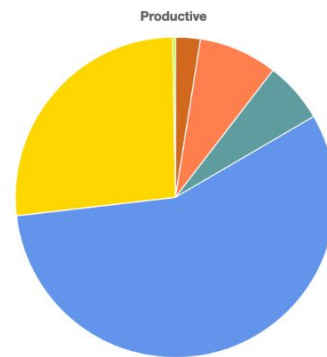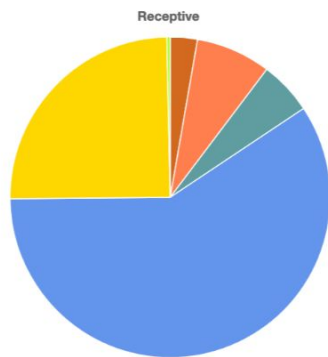
Receptive

Morpheme Category	Receptive
Binding morpheme	990
Inflectional ending	2739
Derivational prefix	1939
Root morpheme	21444

Productive

Morpheme Category	Productive
Binding morpheme	138
Inflectional ending	436
Derivational prefix	338
Root morpheme	3108

Word

Derivational complexity

Item	A1	A2	B1	B2	C1	Total
word	1369	2689	4518	4440	3211	16230
morpheme	2457	5727	11318	11732	9292	40534
morpheme/word	1.79	2.13	2.51	2.64	2.89	2.50

Automatic tools for

- ...detection of morpheme boundaries
- ...labeling morphemes for their types
- ...labeling words for their word formation type
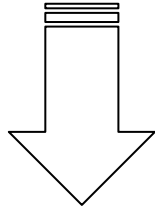
- ...with the ultimate aim/hope to add this type of analysis to automatic pipelines, primarily to Sparv

link to this presentation:

https://tinyurl.com/tmszyfa8