

P17-0716:1

Graded Word Family resource for L2 Swedish

GR4L2 workshop, online Louvain-la-Neuve, Belgium (7 Dec 2021)
Elena Volodina¹, Therese Lindström Tiedemann*, Yousuf Ali Mohammed*¹

*=University of Helsinki, ¹= University of Gothenburg



UNIVERSITY OF GOTHENBURG

SPRÅKBANKENTEXT



UNIVERSITY OF HELSINKI

“L2profiles”

Development of Lexical and Grammatical Competences in Immigrant Swedish

Project financed by

Riksbankens jubileumsfond (2018–2020 (2021))

P17-0716:1

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

Collaboration:

Swedish language, University of Gothenburg
“Nordica”, Dept of Finnish, Finnougric and
Scandinavian Studies, University of Helsinki

PI: Elena Volodina, Gothenburg

Homepage:

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



What is a word family?

Lexical units as the main entry

- word forms → running word forms (walked)
- lemma → base form (walk)
- lemgram → base form + POS (walk, noun)
- flemma → homonymous forms (walk, noun & walk, verb)
- word family → word grouped around the same root (walk, walk-through)

Presuppose different assumptions on vocabulary acquisition. Some discussion see in

Brown, Dale. 2018. Examining the word family through word lists. *Vocabulary Learning and Instruction* 7, no. 1: 51-65.

Stoeckel, Tim, Tomoko Ishii, and Phil Bennett. 2020. Is the lemma more appropriate than the flemma as a word counting unit? *Applied Linguistics* 41, no. 4 (2020): 601-606.

		Word families		
word forms				
lemgram (inflectional paradigms)	2	develop develops developed developing	wood wood's woods wooded	bright brighter brightest
word formation (affixation)	3	developable undevelopable developer(s) undeveloped	woody woodiest woodier woodiness	brightly brightish brightness
	4	development(s) developmental developmentally		
	5	developmentwise semideveloped antidevelopment	wooden	brighten
	6	redevelop predevelopment	anti-wood	

Bauer & Nation
(1993)

- Inclusion of a related form of a word within a **word family** depends on criteria involving *frequency, regularity, productivity and predictability*. These criteria are applied to English *affixes* so that the *inflectional affixes* and the most *useful derivational affixes* are arranged into a **graded** set of seven levels.

Bauer and Nation (1993) [L2 acquisition studies]

Coxhead (1998)

Academic Word List

The following is a list of the words from the Academic Word List that are highlighted in this dictionary. Words shown in **bold** are one of the 'parent words'.

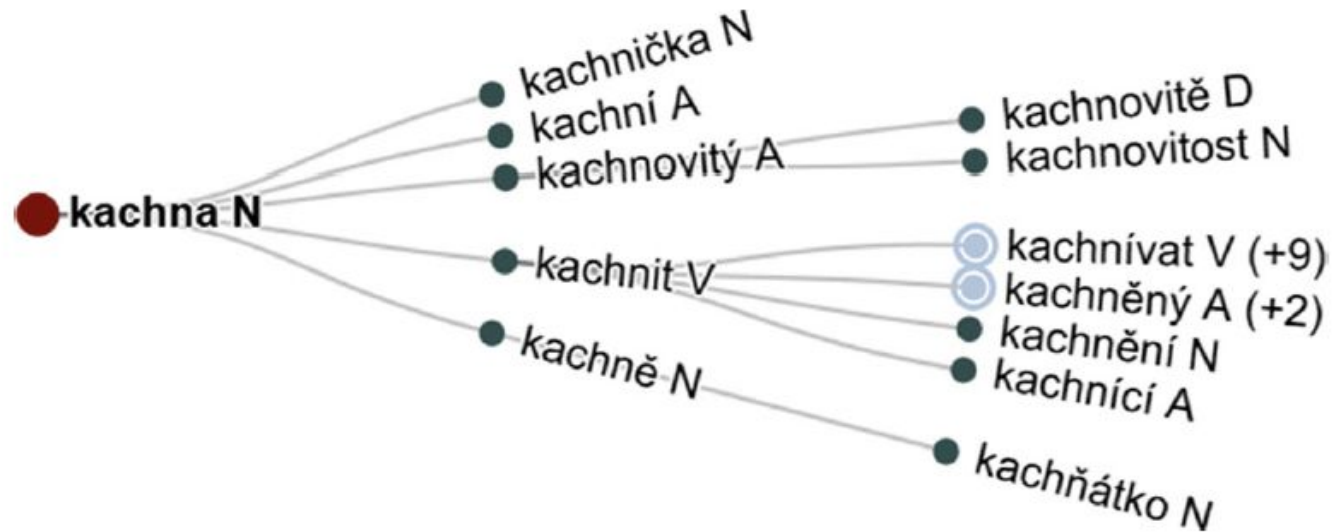
abandon¹*v*
abandoned *adj*
abnormal *adj*
abstract¹*adj*
abstract²*n*
abstract³*v*
abstraction *n*
academic¹ *adj*
academic² *n*
academy *n*
access¹*n*

adjacent *adj*
adjust *v*
adjustment *n*
administration *n*
administrative *adj*
adult¹*n*
adult²*adj*
advocacy *n*
advocate¹*v*
advocate²*n*
affect *v*

analogy *n*
analyse *v*
analysis *n*
analyst *n*
analytical *adj*
analyze *v*
annual¹*adj*
anticipate *v*
anticipation *n*
apparent *adj*
append *v*

assist¹*v*
assistance *n*
assume *v*
assuming *conj*
assumption *n*
assurance *n*
assure *v*
attach *v*
attachment *n*
attain *v*
attainment *n*

Word formation networks: Polish, Czech, Spanish, French



Related resources (a few examples)

- CELEX (Dutch, German, English) lexicons based on L1 corpora
 - <https://catalog ldc.upenn.edu/LDC96L14>
 - contain derivational and compositional structure, inflectional paradigms (among others)
 - protected by paywall
- DErivBase (German, Croatian) - resource with morphological families
 - <https://www.ims.uni-stuttgart.de/forschung/ressourcen/lexika/derivbase/>
 - based on CELEX
- Word formation networks (Polish, Czech, French, Spanish)
 - Lango et al. (2021); DeriNet (lexical network for Czech)
- Word family for Spanish (L2 context)
 - Morin (2006) <https://www.jstor.org/stable/pdf/20063269.pdf>

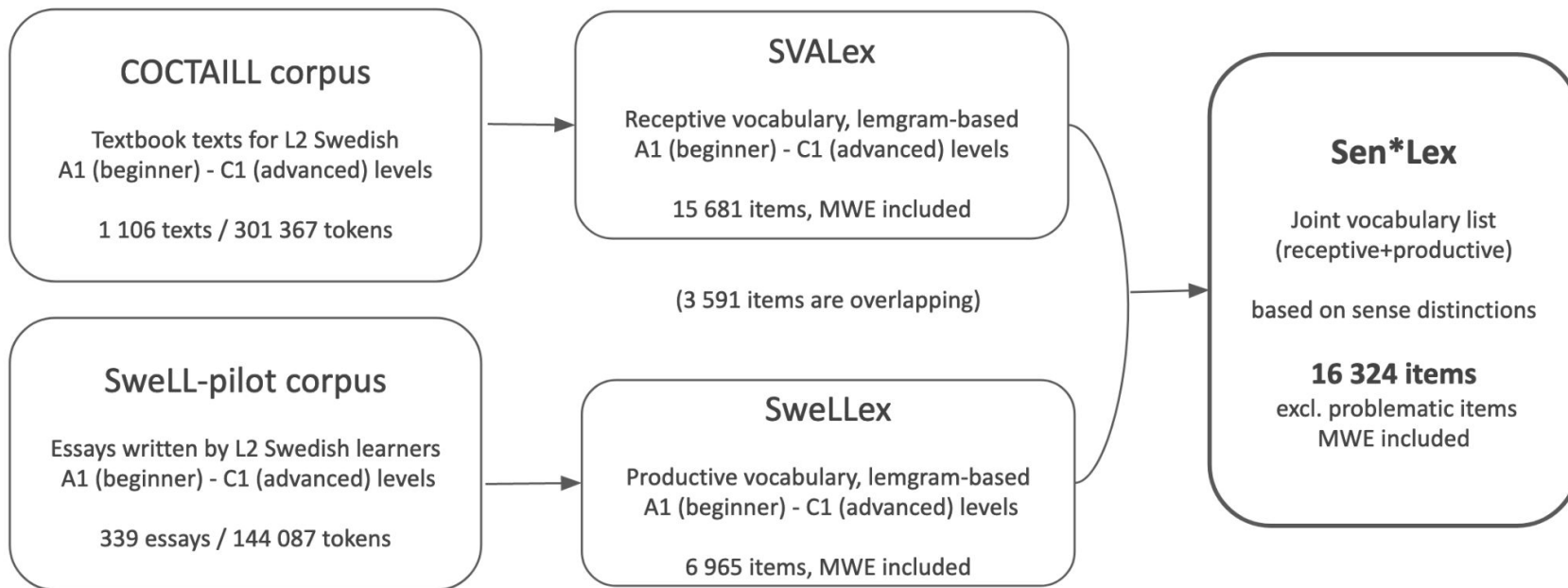
Our definition

Word family is a group of words sharing

- (1) the same root/base, including potential variations of that (e.g. go, bygone), compounds (e.g. light, light-hearted), derived forms (e.g. teach, teacher) and multi-word expressions (e.g. day off, day by day)
- (2) the same meaning (**X** sunlight, light-hearted)

CoDeRooMor dataset:

Compounding, Derivation, Root, Morphology ...and more



Elena Volodina, Yousuf Ali Mohammed, and Therese Lindström Tiedemann. (2021). CoDeRooMor: A new dataset for non-inflectional morphology studies of Swedish. *NoDaLiDa 2021* : Linköping Electronic University Press, Vol.178. [[pdf](#)] [[video-presentation](#)]

Lemgram view

Lemgram	Sense	POS	Analysis	Segment.	Pattern	RealRoot	WordForm	CEFR
adekvat..av.1	adekvat..1	JJ	p:ad r:ekv s:at	ad-ekv-at	p:r:s	rr:adel	derivation	C1
adla..vb.1	adla..1	PC	r:adl s:a	adl-a	r:s		derivation	B2
adel..nn.1	adel..1	NN	r:adel	adel	r		root_lexeme	B1
adelsman..nn.1	adelsman..1	NN	r:adel f:s r:man	adel-s-man	r:f:r		compound	B1
adjektiv..nn.1	adjektiv..1	NN	p:ad r:jekt s:iv	ad-jekt-iv	p:r:s		derivation	A2
adjö..in.1	adjö..1	IN	r:adjö	adjö	r		root_lexeme	A2

“Morpheme family” view

Morpheme	Identifier	Category	Frequency	Examples
a	s	suffix	1 605	leverera, lugna_sig, meritera, narkotika, pumpa, rasa
er	s	suffix	577	abdikera, intrigera, politiker, kritiker, motivera, tekniker
tid	r	root	128	arbetstid, nutid, skoltid, livstid, dåtid, deltid
ny	r	root	46	nyinköpt, nykocht, nykomling, nyligen, nymodighet, Nynäshamn
o	p	prefix	240	olaglig, olämplig, olik, olika, olikhet, oljud
re	p	prefix	105	reaktionstid, rebell, rebellisk, recensent, recensera, recension
s	f	infix*	803	fredstid, landsfader, riksbank, tvångsgift, landsdel, riksdag
o	f	infix*	76	vilopaus, sagobok, sannolik, sociolog, vilorum, typografi

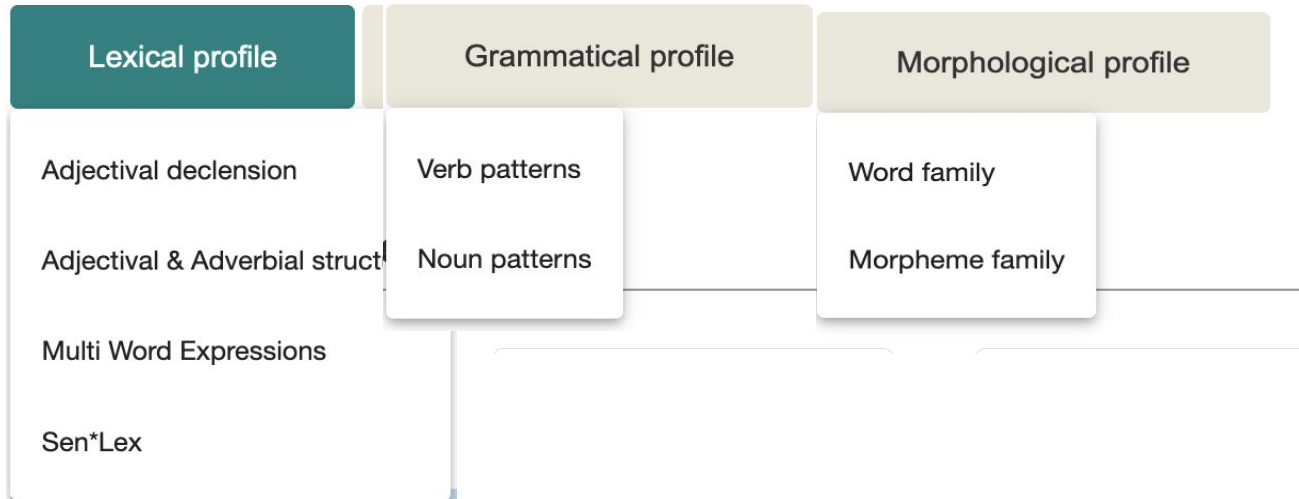
A few facts

- 4 429 word families
- smallest: 1 member (e.g. **adren**alin)
- biggest: 313 members (e.g. **ut**bildning)
- most numerous: function words consisting of root lexemes (e.g. preposition **ut**)

Total rows : 22553 Unique roots/level: A1: 849, A2: 1527, B1: 2168, B2: 2355, C1: 1870,

Our major principles

- research oriented, in addition to teacher- & learner focus
- descriptive, transparent, empirically based, provides access to all corpus hits
- splits receptive and productive knowledge
- statistic analysis and graphical representation





Language Acquisition Reusing **Korp**

Morphological profile – word family

Lexical profile

Grammatical profile

Morphological profile

Swedish Word Family

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

Select root

lär

Search word

WordClass

Saldo WordClass

WordFormation

A1

A2

B1

B2

C1

Table View

Graphical View

Statistical View

Download

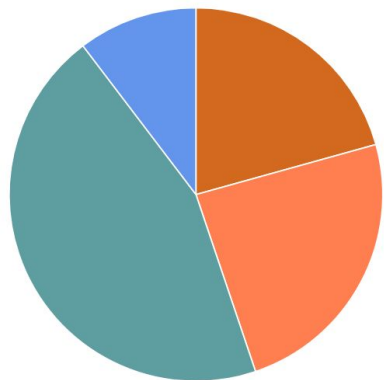
Clear All

Only first occurrence

Total rows: 29

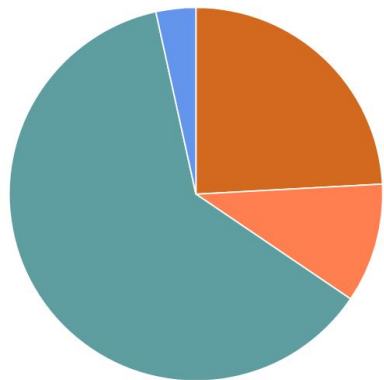
Root ↑↓	CEFR ↑↓	Lemgram ↑↓	Sense ↑↓	WordClass ↑↓	Saldo WordClass ↑↓	WordFormation ↑↓	Receptive ↑↓	Productive ↑↓
lär	A1	lärare..nn.1	lärare..1	Noun (NN)	noun (nn)	Derivation	34	21
lär	A2	svensklärare..nn.1	svensklärare..1	Noun (NN)	noun (nn)	Compound	2	0
lär	A2	lära..vb.1	lära..1	Verb (VB)	verb (vb)	Root lexeme	26	23
lär	A2	lära..vb.1	lära..3	Verb (VB)	verb (vb)	Root lexeme	26	23

WordFormation



■ Lexicalized form
 ■ Derivation
 ■ Compound
 ■ Root lexeme

WordClass



■ Verb (VB)
 ■ Adjective (JJ)
 ■ Noun (NN)
 ■ Participle (PC)

Word family *lär-* 'learn'

Language Acquisition Reusing **Korp**

Lexical profile
Grammatical profile
Morphological profile

Swedish Word Family

A1
A2
B1
B2
C1

Table View
Graphical View
Statistical View
Download
Clear All

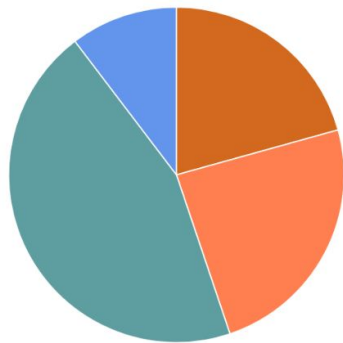
Only first occurrence

Total rows: 29

Root ↕	CEFR ↕	Lemgram ↕	Sense ↕	WordClass ↕	Saldo WordClass ↕	WordFormation ↕	Receptive ↕	Productive ↕
lär	A1	lärare..n.1	lärare..1	Noun (NN)	noun (nn)	Derivation	34	21
lär	A2	svensklärare..n.1	svensklärare..1	Noun (NN)	noun (nn)	Compound	2	0
lär	A2	lära..vb.1	lära..1	Verb (VB)	verb (vb)	Root lexeme	26	23
lär	A2	lära..vb.1	lära..3	Verb (VB)	verb (vb)	Root lexeme	26	23

WordFormation

Mainly compounds but never in production... as 1st occurrence

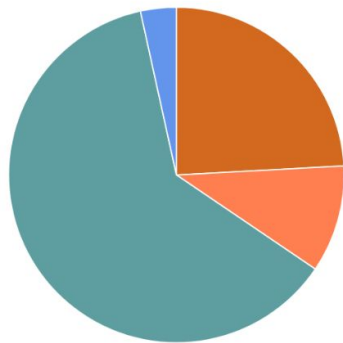


Lexicalized form Derivation Compound Root lexeme

WordClass

Receptive:
mainly nouns

Productive:
mainly verbs



Verb (VB) Adjective (JJ) Noun (NN) Participle (PC)

Word family *lär-* 'learn'

WordFormation	Receptive	Productive
Compound	13	0
Derivation	7	2
Lexicalized form	5	2
Root lexeme	3	2

Number of triplets:
Lemgram + sense + POS

WordClass	Receptive	Productive
Adjective (JJ)	3	0
Noun (NN)	18	2
Participle (PC)	1	0
Verb (VB)	6	4

Select root
lär

Search word

WordClass

Saldo WordClass

WordFormation

Compound

A1

A2

B1

B2

C1

Table View

Graphical View

Statistical View

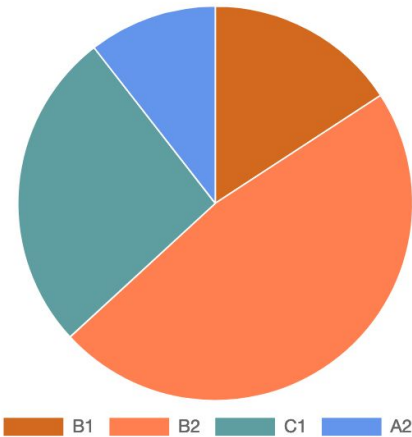
Download

Clear All

Only first occurrence

Total rows: 19

Root ↑	CEFR ↑	Lemgram ↑	Sense ↑	WordClass ↑	Saldo WordClass ↑	WordFormation ↑	Receptive ↑	Productive ↑
lär	C1	språkinläring..nn.1	språkinläring..1	Noun (NN)	noun (nn)	Compound	6	0
lär	B2	lärobok..nn.1	lärobok..1	Noun (NN)	noun (nn)	Compound	6	0
lär	C1	formlära..nn.1	formlära..1	Noun (NN)			3	0
lär	A2	svensklärare..nn.1	svensklärare..1	Noun (NN)			2	0
lär	B2	privatlärare..nn.1	privatlärare..1	Noun (NN)			2	0
lär	B2	svensklärare..nn.1	svensklärare..1	Noun (NN)			2	0
lär	A2	lärobok..nn.1	lärobok..1	Noun (NN)			1	0
lär	C1	självlärd..av.1	självlärd..1	Adjective (JJ)			1	0
lär	C1	lärobok..nn.1	lärobok..1	Noun (NN)			1	0
lär	C1	språklärare..nn.1	språklärare..1	Noun (NN)			1	0



Legend: B1 (blue), B2 (orange), C1 (teal), A2 (brown)

Most common receptive uses per level...

Root ↕	CEFR ↕	Lemgram ↕	Sense ↕	WordClass ↕	Saldo WordClass ↕	WordFormation ↕	Receptive ↕	Productive ↕
lär	A1	lärare..nn.1	lärare..1	Noun (NN)	noun (nn)	Derivation	34	21
lär	A2	lära..vb.1	lära..3	Verb (VB)	verb (vb)	Root lexeme	26	23
lär	A2	lära..vb.1	lära..1	Verb (VB)	verb (vb)	Root lexeme	26	23
lär	A2	lära_sig..vbm.1	lära_sig..1	Verb (VB)	multiword verb (vbm)	Lexicalized form	23	15
lär	C1	språkinläring..nn.1	språkinläring..1	Noun (NN)	noun (nn)	Compound	6	0
lär	C1	formlära..nn.1	formlära..1	Noun (NN)	noun (nn)	Compound	3	0
lär	C1	lära_ut..vbm.1	lära_ut..1	Verb (VB)	multiword verb (vbm)	Lexicalized form	3	0
lär	B2	lärdom..nn.1	lärdom..1	Noun (NN)	noun (nn)	Derivation	3	0

Most common words on A1 and A2 level, in **both receptive and productive**.

Derivation, root lexeme (A1–A2)
Nouns and verbs

Root ambiguity

Swedish Word Family

Select root: Search word: WordClass: Saldo WordClass: WordFormation:

A1 A2 B1 B2 C1 C2

Only first occurrence

Total rows: 16

Root	CEFR	Lemgram	Sense	WordClass	Saldo WordClass	WordFormation	Receptive	Productive
men	A1	men..kn.1	men..1	Conjunction (KN)	Conjunction (kn)	Root lexeme	29.74 (77)	31.74 (13)
men	A1	mena..vb.1	mena..1	Verb (VB)	Verb (vb)	Derivation	0.39 (1)	0.00 (0)
men	A1	promenad..nn.1	promenad..1	Noun (NN)	Noun (nn)	Derivation	0.77 (2)	0.00 (0)
men	A1	promenera..vb.1	promenera..1	Verb (VB)	Verb (vb)	Derivation	1.16 (3)	7.32 (3)
men	A2	sakta_men_säkert..abm.1	sakta_men_säkert..1	Adverb (AB)	Multiword adverb (abm)	Lexicalized form	0.29 (2)	0.00 (0)
men	A2	gemensam..av.1	gemensam..1	Adjective (JJ)	Adjective (av)	Derivation	0.15 (1)	0.00 (0)

but

mean, v

walk

joint, adj

To sum up

- research: through the user interface for empirical analyses
- pedagogical contexts: teaching materials preparation, in classrooms
- ICALL: exercise and test development
- NLP: use as a feature for linguistic complexity measures
- typological studies
- cognitive studies
- ...

Future plans

- Fixing bugs and documentation before the release
- Disambiguation within word families
- Analysis of learning paths (correlating receptive and productive knowledge)
- Exercise/test generation
- ...

Live demo, if we have time...



Thank you!
Comments? Questions?

(or [mailto: elena dot volodina at svenska dot gu dot se](mailto:elena.dot.volodina@svenska.gu.se))