



P17-0716:1

L2Profiles

# Morfologisk profilering av svenska som andraspråk

Språkstrukturseminariet, Göteborgs universitet  
(28 jan. 2021)

Therese Lindström Tiedemann\*, Beatrice Silén\*, Stellan Petersson<sup>1</sup>, Maisa Lauriala\*  
\*=Helsingfors universitet, <sup>1</sup> = Göteborgs universitet



UNIVERSITY OF GOTHENBURG

SPRÅKBANKENTEXT



UNIVERSITY OF HELSINKI

1

## “L2profiler”

Utveckling av lexikala och grammatiska kompetenser i invandrarsvenska



UNIVERSITY OF GOTHENBURG

SPRÅKBANKENTEXT



RIKSBANKENS  
JUBILEUMSFOND  
FRÄMJAR HUMANIORA  
OCH SAMHÄLLSVETENSKAP

### Projekt finansierat av:

Riksbankens jubileumsfond (2018–2020 (2021)) P17-0716:1 (gemensam ansökan Elena & Therese)

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

### Samarbete:

Svenska språket, Göteborgs universitet och Nordica, Helsingfors universitet

**Projektledare:** Elena Volodina, Göteborgs universitet

### Hemsida:

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



UNIVERSITY OF HELSINKI

2

# “Profilering”

Data → “Typers” frekvens →  
översiktsbild över utvecklingen  
(Granfelt & Ågren 2014)



Profiling →  
CEFR referensnivådeskriptorer:  
*Profile Deutsch; English Profile*

Jfr även *Lexical Frequency Profile*  
(Laufer & Nation 1995, Laufer  
2005)

**Ursprung:** Speech disorder  
research (Crystal, 1976); spreds  
sen till SLA (Clahsen 1985; 2012)

3

## English Vocabulary Profile Online - British English

Search... A1 A2 B1 B2 C1 C2 Select All Search Advanced Search Clear Results

Topic - Select - Part of Speech - Select -

Hide culturally sensitive words  Yes

Category - Select - Grammar - Select - Usage - Select -

Prefix - Select - Suffix - Select -

4

| Prefix                       |                     | Suffix     |                |                     |                         |           |           |    |
|------------------------------|---------------------|------------|----------------|---------------------|-------------------------|-----------|-----------|----|
| dis-                         |                     | - Select - |                |                     |                         |           |           |    |
| Results 1 - 20 of 55         |                     |            |                | Sort by:            | Level                   | Ascending | Display # | 20 |
| Base Word                    | Guideword           | Level ^    | Part of Speech | Topic               | Details                 |           |           |    |
| disadvantage                 | MAKING DIFFICULT    | B1         | noun           |                     | <a href="#">Details</a> |           |           |    |
| dislike                      |                     | B1         | verb           | people: personality | <a href="#">Details</a> |           |           |    |
| disagree                     |                     | B1         | verb           | communication       | <a href="#">Details</a> |           |           |    |
| likes and dislikes           |                     | B1         | phrase         | people: personality | <a href="#">Details</a> |           |           |    |
| disabled                     |                     | B1         | adjective      | body and health     | <a href="#">Details</a> |           |           |    |
| disappear                    | NOT SEE             | B1         | verb           |                     | <a href="#">Details</a> |           |           |    |
| disappear                    | NOT FIND            | B1         | verb           |                     | <a href="#">Details</a> |           |           |    |
| disappearance                |                     | B2         | noun           |                     | <a href="#">Details</a> |           |           |    |
| discourage                   | MAKE LESS CONFIDENT | B2         | verb           | communication       | <a href="#">Details</a> |           |           |    |
| discourage sb from doing sth |                     | B2         | phrase         | communication       | <a href="#">Details</a> |           |           |    |
| disapprove                   |                     | B2         | verb           | people: personality | <a href="#">Details</a> |           |           |    |

5

## I dag

- Varför morfologi?
- Annotering
- Exempel morfologisk annotering
- Samstämmighet & Homografer
- Utveckling?

8

# Varför morfologi?

9

## Varför morfologi?

- L1 och L2 forskning – ordförrådet viktigt för språkförståelse (Mochizuki & Aizawa 2000, Schmitt & Meara 1997, Amirjalili and Akbar Jabbari, 2018: 3; Sukying, 2020)
- Ordförrådsforskning (Nagy et al 2013:3) – morfologi och ordfamiljer (Bauer m.fl. 1993))
- Inbyggt i vokabulärtest (t.ex. *Word part levels test* Webb & Sasao 2017; *Vocabulary levels test* Nation 1983, Webb m.fl. 2017; *LFP* Laufer & Nation 1995)

10

## Varför morfologi?

- L1 och L2 forskning – ordförrådet viktigt för förståelse (Mochizuki & Aizawa 2000, Schmitt & Meara 1997, Amirjalili and Akbar Jabbari, 2018: 3; Sukying, 2020)
- Ordförrådsforskning (Nagy et al 2013:3) – morfologi och ordfamiljer (Bauer m.fl. 1993))
- Inbyggt i vokabulärtest (t.ex. *Word part levels test* Webb & Sasao 2017; *Vocabulary levels test* Nation 1983, Webb m.fl. 2017; *LFP* Laufer & Nation 1995)

→ **ordfamiljer** vore intressant – utvecklingen av L2

→ **ordfamilj** och morfeminformation behövs för att utveckla test för svenska såsom internationella test

11

## Morfem?

- Morfem – minsta betydelsebärande enheten
- Vilken är den minsta betydelsebärande enheten i orden?
- Hur fastställer vi det?
  - Jämföra med andra ord
- Är det samma för L1-talare och L2-talare?

12

## Morfem?

- Morfem – minsta betydelsebärande enheten
- Vilken är den minsta betydelsebärande enheten i orden?
- Hur fastställer vi det?
  - Jämföra med andra ord
- Är det samma för L1-talare och L2-talare?

Vi återkommer strax till hur vi gjort när vi analyserat.

13

## Annoteringsarbetet

14

## Data

16324 Lemgram + POS + betydelse extraherade ur:

- **Coctail-korpusen** (Volodina et al., 2014) – kursböcker Svenska som andraspråk (sverigesvenska). (CEFR-nivåer)
  - 594 287 token som länkats till en nivå (A1–C1)
- **SweLL-pilot-korpusen** (Volodina et al., 2016) – inlärtartexter
  - 155282 token (A1–C2)

15

|                    | <b>A1</b> | <b>A2</b> | <b>B1</b> | <b>B2</b> | <b>C1</b> | <b>C2</b> |         |
|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|---------|
| <b>Coctail</b>     | 34 464    | 80 677    | 147 363   | 161 740   | 170 043   | -         | 594 287 |
| <b>SweLL-pilot</b> | 5 172     | 26 200    | 23 921    | 45 182    | 51 549    | 3 258     | 155 282 |

16

## Annoteringens utveckling

1. Legato – annoteringsverktyg – första runda under hösten 2019 (Beatrice och Stellan)
  - Mål: annotera med ett L2 perspektiv
  - Resultat: Jämförelse visade på många fall med olika resonemang
2. Ny analys baserad på **Svensk ordbok**
  - **Samarbete** med redaktionen för *Svensk ordbok* utgiven av Svenska Akademien: deras morfemanalys (utan morfemkategori) direkt inne i Legato
3. Parallellannotering: samma 100 x 4, jämförde (per 100), oannoterade
  - (automatisk jämförelse (Yousuf
4. Resten uppdelat på 3 annoterare – jämförelse på morfemnivå
  - Beatrice och Stellan – tillgång till sina tidigare annoteringar

18

## Morfemanalys

- Synkron i första hand – jämför med andra ord i språket
- Bör stämma med SO och SAG
- Om dessa inte stämmer överens → diskussion → dokumentation

19



## Morfemanalys – etymologi?

- Synkron i första hand
- Bör stämma med SO och SAG
- Om dessa inte stämmer överens → diskussion → dokumentation
- Etymologi – tas i bruk för att upptäcka möjliga paralleller
  - *Frilans*: SO – “sedan 1945 av eng. freelance med samma betydelse, eg. 'fri lans', urspr. om legoknekt; jfr [lans](#)” → även *lansera*: SO – “sedan 1892 av fra. lancer med samma betydelse, eg. 'slunga lans'; jfr [lans](#)”
- Etymologi kan också användas som hjälp: rot eller affix?
- Synkron morfemindelning kan vara annorlunda (jfr reanalys, grammatikalisering): *hamburger* > *burger*
- Svensk morfemindelning kan vara annorlunda än långgivande språk:
  - *Marshmallow*: SO *marsh·mallow·en* – L2P *marshmallow* (rot)

20

## Samarbete med redaktionen för SO och SAOL

- SO och SAOL inne i Legato – STOR HJÄLP!
  - Morfemkategori saknas – vi lägger till
  - Fogemorfem särskiljs inte – vi markerar detta som eget morfem
  - Måste ändå titta i SO och SAOL på svenska.se ofta
  - Även SAOB, SAG, KORP, SALDO



SVENSKA AKADEMIENS  
ORDBÖCKER



Lexicographic Annotation Tool (LEGATO)

21

## Morfologisk analys i SO: ex *allomfattande* – Vilken ordbildningstyp? Vilken morfemkategori?

**SO**

tryckår: 2009

**all`omfattande** adjektiv, ingen böjning

ORDLED: all|om|fatt-ande

- som omfattar allt och alla

[Visa mer](#)

**SO**

tryckår: 2009

**omfattande** [ãm`-] adjektiv, ingen böjning

ORDLED: om|fatt-ande

- som har stort omfång

JFR avsevärd

[Visa mer](#)

22

## Annotering i Legato

(utv. David Alfter m.fl., vidareutv. Yousuf Ali Mohammed)



### Lexicographic Annotation Tool (LEGATO)

Guidelines Skipped items **98** Search Filter External links

Quick jump to: A B C D E F G H I J K L M N O P Q R S T U V W X Y Z A Å Ö

Jump to:  Jump

1 - 19314

Current task: MORPHOLOGY

Progress: 12739/16314

SALDO lemgram

språkvårdare..nn.1

Saldo POS

noun (nn)

Part-of-Speech

Noun (NN)

CEFR level

C1

Saldo sense: språkvårdare..1  
Saldo primary descriptor: språkvård..1  
Saldo secondary descriptor: PRIM..1

#### Examples:

Svenska akademien ger ut ordlista och grammatik och många stora tidningar har egna \*\* språkvårdare \*\* .  
I regeringskansliet sitter \*\* språkvårdare \*\* som ska se till att myndigheterna inte krånglar till språket för mycket .

Legato (v, 1): Alfter, D., Lindström Tiedemann, T & Volodina, E. (2019)

23

|  |  |  |  |
|--|--|--|--|
| <b>Previous morphology1:</b><br><div style="border: 1px solid gray; height: 40px; width: 100%;"></div> <input type="button" value="Copy"/>   | <b>Previous morphology2:</b><br><div style="border: 1px solid gray; height: 40px; width: 100%;"></div> | <b>SO analysis:</b><br><div style="border: 1px solid gray; padding: 5px;">språk=vård-ar-en</div> <input type="button" value="Copy"/> | <b>SAOL analysis:</b><br><div style="border: 1px solid gray; padding: 5px;">språk=vård-are</div> <input type="button" value="Copy"/> |
| <b>Current values:</b><br><div style="border: 1px solid gray; padding: 5px;"> r:språk<br/> r:vård<br/> s:ar<br/> te </div>   |  |  |  |
| <b>Select word form:</b><br><input type="radio"/> ROOT LEXEME<br><input type="radio"/> DERIVATION<br><input checked="" type="radio"/> COMPOUND<br><input type="radio"/> LEXICALIZED FORM<br><input type="radio"/> UNKNOWN<br><input type="radio"/> ABBREVIATION<br><br><input type="checkbox"/> MARK FOR REGISTER<br><input type="checkbox"/> L2 ANALYSIS LIKELY TO DIFFER |  |  |  |
| <input type="button" value="Exit"/> <input type="button" value="Previous"/> <input type="button" value="Skip"/> <input type="button" value="Save &amp; Next"/>   |  |  |  |

24



LÄR språket via Korpus Analys

## Annotering i Legato

### Morfemkategorier

- Rot (r)
- Prefix (p)
- Böjningsändelse (i)
- Avledningssuffix (s)
- Fogemorfem (f)
- ?
  
- rr ("real root") - t.ex. *kaotisk* ->  
r:kaot  
rr:kaos  
s:isk

### Lexicographic Annotation Tool (LEGATO)

#### Ordbildningstyp

- Rotlexem, t.ex. *läsa, saga*
- Avledning, t.ex. *orolig, teknik*
- Sammansättning, t.ex. *sysselsätta, utanför*
- Lexikaliserad
  - Flerordsenheter, t.ex. *i alla fall*
  - Namn som t.ex. *Aftonbladet, Svenska Akademien*
  - Ord som oftast eller alltid förekommer i pluralis, t.ex. *sötsaker*
- Förkortning, t.ex. *AB, jfr* -> analyseras inte
- Unknown, t.ex. personnamn

25

## Exempel och diskussion

26

### Samstämmig analys – vissa principer

- Böjningsändelse eller Avledningsändelse?
- Definitioner
- Vi behöver principer och riktlinjer

27

## Samstämmig analys – vissa principer

- Böjningsändelse eller Avledningsändelse?
- Definitioner
- Vi behöver principer och riktlinjer
- Ändras i andra böjningsformer, inte del av avledningsstammen → böjningsändelse
  - *pojke*, *pojkar*, *pojkskola*, *pojktig*
- Blir del av böjningsstammen och ev. även avledningsstammen → avledningsändelse
  - *minne*, *minnen*; *minnesstund* jfr även *minnas*

28

## -e i slutet av substantiv

|               | Ordböjning               | Sammansättning             | Avledning |                    |
|---------------|--------------------------|----------------------------|-----------|--------------------|
| <b>Typ 1:</b> | Pojk-e, -ar              | pojkskola                  | pojktig   | böjningsändelse    |
| <b>Typ 2:</b> | Rik-e, -en               | Riksdag<br>Riksarkiv       |           | avledningssuffix   |
|               | Minn-e, -en<br>(minn-as) | Minnesalbum<br>Minnesstund |           |                    |
| <b>Typ 3:</b> | Aktie, -r                | Aktieägare                 |           | -e är del av roten |

29

## A som morfem

|               |                   | Ordböjning   | Samman-sättning          | Avledning              | Morfemkategor<br>i     |
|---------------|-------------------|--|--------------------------|------------------------|------------------------|
| <b>Verb</b>   | läs <b>a</b>      | läs!, läser, läste,  | läs-utmaning             | läs <b>are</b>         | Böjningsändelse        |
|               | kräk <b>as</b>    | kräks!, kräks, kräktes                                     |                          |                        | Böjningsändelse        |
|               | tal <b>a</b>      | tal <b>a</b> !, tal <b>ar</b> , tal <b>ade</b>             |                          | tal <b>are</b> , (tal) | Avlednings-<br>ändelse |
|               | planer <b>a</b>   | planer <b>a</b> !, planer <b>ar</b> ,<br>planer <b>ade</b> |                          | planering              | Avlednings-<br>ändelse |
| <b>Subst.</b> | kvinn <b>a</b>    | kvinn <b>an</b> , kvinnor                                  | kvinn <b>o</b> -förening | kvinn <b>l</b> ig      | Böjningsändelse        |
|               | kung <b>a</b> hus |  |                          |                        | Fogemorfem             |

30

## Vad är rot, böjningssuffix och avledningssuffix?

SAG:

“morfem som ensamt eller tillsammans med böjningssuffix kan utgöra ett självständigt ord: *på, och, att, trä, jämn, böj*. Rotmorfemet kan bilda avledningar tillsammans med avledningssuffix (*jämnh*) och avledningsprefix (*ojämn*) eller sammansättningar tillsammans med andra ord (*jämntjock*).”

31

## Sammansättningar eller avledningar?

|              | SO              | SAG   |                      | Vår analys                     |
|--------------|-----------------|---|----------------------|--------------------------------|
| åtala        | å  tal•a        | ?   | sammansättning(?)    | r:å<br>r:tal<br>s:a            |
| ogilla       | o  gill•a       | prefix nya<br>substantiv<br><i>otur</i>     | avledning! (jfr SAG) | p:o<br>r:gill<br>s:a           |
| ombud        | om  bud         | ?   | sammansättning?      | (p?)/r:om<br>r:bud             |
| missförstånd | miss  för•stånd | prefix nya<br>substantiv:<br><i>missjud</i> | sammansättning?      | (p?)r:miss<br>p:för<br>r:stånd |

32

## Inlånade ord: Ordbildning?

|         | SO                       | Etymologi (SO)   | Vår analys                                 |
|---------|--------------------------|--|--|
| affär   | af•fär<br>→ avledning?   | sedan ca 1890 av<br><b>fra. affaire</b> 'an-<br>gelägenhet, göro-<br>mål', till uttrycket <b>à<br/>faire</b> '(som är) att<br>göra'  | p:af<br>r:fär<br><br>eller<br>r:affär?     |
| alfabet | alfa•bet<br>→ avledning? | sedan 1753 av <b>lat.<br/>alphabe'tum</b> med<br>samma betydelse;<br>till <b>al'pha 'a'</b> och<br><b>be'ta 'b'</b> , namn på<br>de två första bok-<br>stäverna i det grek.<br>alfabetet | ?:alfa<br>?:bet<br><br>eller<br>r:alfabet? |

33

## Vilka delar består avledningen av?

|          | SO         | Etymologi (SO/SAOB)  | Vår analys   |
|----------|------------|--|--|
| Religion | re•lig-ion | sedan 1539 av <b>lat. relig'io</b> 'förpliktelse; samvete; gudadyrkan', till <b>religa're</b> 'binda fast'; jfr <a href="#">reliabel</a>   | p:re<br>r:lig<br>s:ion<br><br>Eller r:relig?                             |
| Kaotisk  | kaot•isk   | [jfr t. chaotisch, eng. chaotic, fr. chaotique; bildat till <a href="#">KAOS</a> efter mönster av sådana avledn. till gr. ord som <a href="#">EROTISK</a> , <a href="#">HYPNOTISK</a> m. fl.]                          | r:kaot<br>s:isk<br>rr:kaos   |
| Refräng  | refr•äng   | sedan 1791; tänka på refrängen 1847 av fra. refrain 'återsvall; refräng'; nära besl. med <a href="#">fraktion</a><br>SAOB: av mlat. refrangere, bryta tillbaka, bryta på nytt, av lat. re- (se RE-) o. frangere, bryta | r:refr<br>s:äng (jfr gratäng, maräng)<br><br>Eller???<br>p:re<br>r:fräng |
| Semester | sem•ester  | sedan 1756 via ty. av lat. seme'stris 'som varar sex månader', till sex' 'sex' och me'nsis 'månad'; urspr. i bet. 'halvår; termin'   | r:se<br>r:mest<br>s:er   |

34

## Analys av ord med *-graf*, *-krat*, *-log* och *-nom*

- Vi ser *-graf*, *-krat*, *-log* och *-nom* som rötter som kan specificeras av olika prefix.

|          |       |       |
|----------|-------|-------|
| Exempel: | p:eko | p:agr |
|          | r:nom | f:o   |
|          | (s:i) | r:nom |
|          |       | (s:i) |

|                |                  |               |
|----------------|------------------|---------------|
| foto-graf-(i)  | tele-graf-(i)    | etno-graf-(i) |
| demo-krat-(i)  | arist-o-krat-(i) | byrå-krat-(i) |
| psyk-o-log-(i) | geo-log-(i)      | bio-log-(i)   |

35



## Analys av ord med *-graf*, *-krat*, *-log* och *-nom*

“[Gr]af känner vi igen från många andra ord: *fotograf*, *etnograf*, *stenograf* m.fl. Det verkar finnas en vag betydelselikhet, *graf* tycks beteckna en person som sysslar med (ofta beskriver) något som framgår av det första ledet. Men om vi vill etablera *graf* som morfem i *geograf* återstår det att kontrollera om vi känner igen resten av ordet, *geo-*. Det finns åtminstone i ordet *geolog* (där i sin tur det senare ledet *-log* känns igen från många andra ord), och de flesta svenska språkbrukare vet nog åtminstone på ett allmänt plan att *geo-* i båda fallen har med jorden att göra. Vi kan fullfölja analysen genom att **fastslå att *graf* bör betraktas som ett avledningssuffix** och *geo* som ett bundet rotmorfem.”

(Malmgren 1994: 29, fetstil L2P)

36

## Inte alltid samma princip: *bio*, *eko* och *tele*

Trots att *eko* i *ekologi* och *ekonomi* ses som prefix kan de bli rot i andra ord.

|                          |                                     |  |
|--------------------------|-------------------------------------|--|
| <i>ekoturist</i>         | SAOL: eko tur-ist<br>SO: tur·ist·en | Sammansättning:<br>r: <i>eko</i><br>avl: <i>turist</i>       |
| <i>bio</i>               |                                     | Rot  |
| <i>biogas</i>            | SO: bio gas·en                      | Sammansättning:<br>r: <i>bio</i><br>r: <i>gas</i>            |
| <i>telekommunikation</i> | SO: tele kom·mun·ik·at·ion·en       | Sammansättning:<br>r: <i>tele</i><br>s: <i>kommunikation</i> |
| <i>telefon</i>           | SO: tele·fon·en                     | Avledning:<br>p:tele<br>r:fon                                |

Jfr  
*lexicalisation*,  
*clipping* &  
*syntacticization*  
(Brinton &  
Traugott  
2005):  
bundna delar  
> nya lexem,  
t.ex. *omnibus*  
> *bus(s)*,  
*automobil* >  
*bil* / *auto*

37

## Substantiv som slutar på *i*

- T.ex. *monarki, akademi, ironi, teori*

|            |            |          |
|------------|------------|----------|
| SO         | akademi·er | teori·er |
| Vår analys | r:akadem   | r:teor   |
|            | s:i        | s:i      |

- SO särskiljer inte *-i*: har vanligtvis inte morfem på en bokstav
- Vi vill koppla ord till varandra för att kunna hitta ordfamiljer

|         |            |             |
|---------|------------|-------------|
| Jämför: | akadem·isk | teor·et·isk |
|         | r:akadem   | r:teor      |
|         | s:isk      | s:et        |
|         |            | s:isk       |

38

## Länder, nationalitet...

### *Italien*

ROT: *itali*,  
AVL.SUFFIX: *en* (böjning? Jfr Bakken 2002)

### *Italienare* (nationalitet, subst.)

ROT: *itali*,  
AVL.SUFFIX: *en*,  
AVL.SUFFIX: *ar*,  
BÖJNINGSSUFFIX: *e*

### *Italiensk* (nationalitet, adj.)

ROT: *itali*,  
AVL.SUFFIX: *en*,  
AVL.SUFFIX: *sk*

## SO

tryckår: 2009

**italie`nare**  substantiv ~n äv. vard.

*italienarn*, plur. ~, best. plur. *italienarna*

ORDLED: itali·en·ar·en

*feg-is*! När flerstaviga ord på *-isk* med accent 1 utgör stam i avledning med suffixet *-a* bibehålls tonaccenten: *grekisk-a*, *slavisk-a*. Substantiv på *-anska*, *-enska* har också vanligen accent 1 (särskilt i beteckning för språk), men också accent 2 förekommer (särskilt i beteckning för kvinna): *itali<sup>1</sup>enska* eller *itali<sup>2</sup>enska*, *ja<sup>1</sup>panska* eller *ja<sup>2</sup>panska*.

SAG del 2: Substantiv §20 (s. 39)

41

## Länder, nationalitet...

### Italien

ROT: *itali*,

AVL.SUFFIX: *en* (böjning? Jfr Bakken 2002)

**Italienare** (nationalitet, subst.)

ROT: *itali*,

AVL.SUFFIX: *en*,

AVL.SUFFIX: *ar*,

BÖJNINGSSUFFIX: *e*

**Italiensk** (nationalitet, adj.)

ROT: *itali*,

AVL.SUFFIX: *en*,

AVL.SUFFIX: *sk*

**SO**

---

tryckår: 2009

**italie`nare** substantiv ~n äv. vard.  
*italienarn*, plur. ~, best. plur. *italienarna*  
 ORDLED: itali-en-ar-en

*feg-is!* När flerstaviga ord på *-isk* med accent 1 utgör *stam* i avledning med suffixet *-a* bibehålls tonaccenten: *grekisk-a*, *slavisk-a*. Substantiv på *-anska*, *-enska* har också vanligen accent 1 (särskilt i beteckning för språk), men också accent 2 förekommer (särskilt i beteckning för kvinna): *itali<sup>1</sup>enska* eller *itali<sup>2</sup>enska*, *ja<sup>1</sup>panska* eller *ja<sup>2</sup>panska*.

SAG del 2: Substantiv §20 (s. 39)

42

## Sverige, svensk?? Hur får vi ihop ordfamiljen?

|         |   | Analys?   | Ord i Legato (ex)  | Etymologi  |
|---------|---|-----------|--|--|
| Sverige |   | r:sverige |  |  |
| svensk  | SO: (svensk <i>adj</i> ),<br>svensk-en <i>subst</i> . | r:sven    | <i>Svensk, svenskfödd, svenskfödd, svenskhet, svenskurs, svensktalande, sydsvensk, utlandssvensk</i> | fsv. svänsker, motsv. fd. svensk (d. svensk), nor. svensk, fvn. svenskr, søenskr; sannol. i avljudsförh. till <a href="#">SVEAR.</a> |
| svenska | SO: svensk-an   | r:sven    |  |  |
| svensk  |   | r:svensk  | <i>osvensk, rikssvensk, rikssvenska, finlandssvensk, försvenska, försvenska, helsvensk</i>           |  |
| svear   | SO: sve-ar  |           |  |  |

43

|                 |                         |   |   |
|-----------------|-------------------------|---|---|
| norsk           | so: nor-ska             | <i>norsk, norsk, nynorsk</i>  | r:nor                                     |
| norrman         | so: norr mann-en        | <i>norrman</i>  | r:norr (rr: nor?)                         |
| dansk           | so: dan-ska             | <i>danmark, dansk</i>   | r: dan / r:dansk                          |
| finsk           | so: fin-ska             | <i>finland, finlandssvensk, finsk<br/>finsk-ugrisk, finsktalande</i>  | r:fin / r:finsk                           |
| tornedalsfinska | so: torne-dals finsk-an | <i>tornedalsfinska</i>  | r:fin                                     |
| finländare      | so; fin länd-ar-en      | <i>finländare</i>   | r:fin                                     |
| engelsk         | so: engel-ska           | <i>England, engelsk, engelska, engelsman,<br/>engelsklärare, engelskspråkig,<br/>engelsktalande</i>           | r:eng<br>r:engel<br>r:engels<br>r:engelsk |
| fransk          | so: fransk              | <i>fransman, fransson, fransyska<br/>fransk, fransk, franska_revolutionen,<br/>fransktalande, skolfranska</i> | r:frans<br>r:fransk                       |
| italiensk       | so: itali-en-ska        | <i>italien, italienare, italiensk, italiensk</i>  | r:itali                                   |
| spansk          | so: span-ska            | <i>spansk, spanska, spansktalande<br/>Spanjor, Spanien</i>  | r:span<br>r:spanj<br>r:spani              |

44

27 jan 2021

## Sk som avledningssuffix

| Corpus      | A1  | A2  | B1  | B2  | C1  | C2 | Total |
|-------------|-----|-----|-----|-----|-----|----|-------|
| Legato      | 15  | 15  | 20  | 22  | 21  | 0  | 93    |
| Coctail     | 107 | 222 | 423 | 362 | 416 | 0  | 1530  |
| Swell-pilot | 14  | 73  | 39  | 45  | 42  | 3  | 216   |

s -> afrikansk, koreansk, latinsk, lutheransk, luthersk, medicinsk, mexikansk, muslimsk, nigeriansk, nordafrikansk, nordamerikansk, norsk, norsk, nynorsk, ondska, polsk, polsk, professorska, romansk, romersk, Sixtinska\_kapellet, sjuksköterska, skådespelerska, skånsk, sköterska, nederländsk, norrländsk, nyzeeländsk, småländsk, iransk, islamsk, italiensk, italiensk, japansk, karolinsk, Karolinska, kassörska, katolsk, spartansk, svengelska, svensklärare, svägerska, sydafrikansk, sångerska, sömmerska, tibetansk, tornedalsfinska, trojansk, undersköterska, ungersk, urgermansk, urgermanska, viktoriansk, västafrikansk s -> icke-svensk, kanslīsvenska, svensk-tysk, svensk, svensk, svensk, svensk, Svenska\_Akademien, Svenska\_Akademien, Svenska\_Dagbladet, Svenska\_Dagbladet, svenskfödd, svenskfödd, svenskhet, svenskurs, svensktalande, sydsvensk, utlandssvensk, isländsk, thailändsk, utländsk, värmländsk, västerländsk...

45

## Samstämmighet och homografer...

46

### Möten och parallellannotering

- 400 ord annoterades av alla 4 (dvs Therese, Beatrice, Stellan och Maisa)
- Veckomöten – jämfört på morfemnivå för att fånga oenigheter

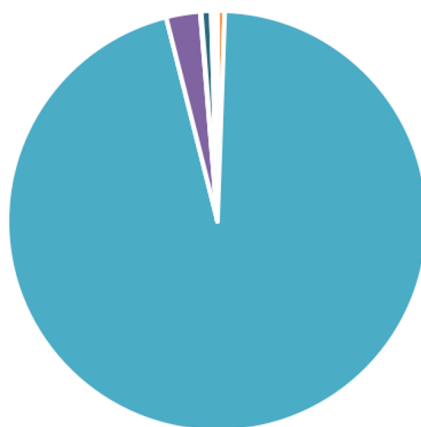
47

## Hur lika annoterar vi?

|  |                                  |              |   |
|--|----------------------------------|--------------|---|
| Enligt automatisk jämförelse               |                                  |              |   |
| Annoterade lika                            | (667 + 782 =) <b>1449 morfem</b> | 30,3 %       | Inkl homografer av samma typ.                                   |
| Annoterade "olika"                         | <del>3122</del> 27 morfem        | <b>0,6 %</b> |   |
| Annoterade av bara en                      | (3122-27=) 3095                  | 64,8 %       |   |
| Annoterade delvis olika                    | (986-782 =) <b>204 morfem</b>    | <b>4,3 %</b> | Homografer men även fel   |
| Annoterade lika av två och inte alls av en | 782 morfem                       |              |   |
| TOTALT                                     | 4775 "morfem"                    | 100 %        | Vissa av dessa är homografer på morfemnivån → morfemen är fler. |

48

### Lika annoterade av alla eller av 2 (plus en blank)



**Homografer:**  
5,3 % av morfemen har sett som >1 morfemtyp av minst 1 annoterare.



49

## Homografer

|   |                  |                    |
|---|------------------|--------------------|
| s | Rot              | Bokstaven <b>s</b> |
|   | Avledningssuffix | Inte all <b>s</b>  |
|   | Böjningssuffix   | and <b>s</b>       |
|   | Fogemorfem       | arbet <b>s</b> dag |

50

## Homografer – deponens = böjning?

|   |                       |                    |
|---|-----------------------|--------------------|
| s | Rot                   | Bokstaven <b>s</b> |
|   | Avledningssuffix      | Inte all <b>s</b>  |
|   | <b>Böjningssuffix</b> | and <b>s</b>       |
|   | Fogemorfem            | arbet <b>s</b> dag |

51

## Homografer – deponens = böjning? (SAG)

|     |                       |                     |
|-----|-----------------------|---------------------|
| -s- | Rot                   | Bokstaven <b>s</b>  |
|     | Avledningssuffix      | Inte all <b>s</b>   |
|     | <b>Böjningssuffix</b> | and <b>as</b>       |
|     | Fogemorfem            | arbets <b>s</b> dag |

SAG del 2: Verb §49

§ 48. **S-verb som deponens.** När verbets *s*-form inte har passiv funktion fungerar den syntaktiskt ungefär som ett intransitivt (eller någon gång: transitivt) verb i *s*-lös form. *S*-formen kallas då deponens.<sup>1</sup>

Den deponentiella *s*-formen är liksom den passiva en böjningsform av verbet: suffixet föregås av verbets övriga böjningssuffix. Funktionellt-semanticiskt är den deponentiella *s*-formen emellertid snarast jämförbar med en avledning: vissa betydelseundergrupper kan urskiljas och produktiviteten är starkt begränsad, dvs. deponentiell *s*-form förekommer bara av vissa verb.

52

## Pågående forskningsfrågor just nu...

- Hur ökar morfemen över CEFR-nivåerna?
- Vi började från: **16 324 ord**
- Några få ord annoteras inte pga fel.
- **4775 “morfem”**

53



## Morfemkategori / nivå – ca 16000 ord

### Morpheme category/CEFR level

| Identifier          | A1   | A2   | B1   | B2   | C1   | Total |
|---------------------|------|------|------|------|------|-------|
| Root (orthographic) | 1606 | 3503 | 6416 | 6696 | 5042 | 23268 |
| Inflectional ending | 239  | 454  | 845  | 848  | 570  | 2958  |
| Suffix              | 438  | 1244 | 2721 | 2744 | 2453 | 9601  |
| Prefix              | 79   | 215  | 613  | 598  | 580  | 2085  |
| Fogemorfem          | 20   | 112  | 260  | 322  | 251  | 965   |
| Unknown             | 12   | 30   | 63   | 45   | 25   | 175   |
| <b>Root (real)</b>  | 38   | 82   | 161  | 174  | 130  | 585   |

26 jan. 2021

54

## Morfemtyper / nivå (baserat på de annoterade lexemen)

### Morpheme types according to morpheme categories

*The number outside the parenthesis is the total number of morpheme types develop according to morpheme categories at a particular CEFR level and the number inside the parenthesis is the new morpheme types develop according to morpheme categories at that level.*

| Identifier          | A1        | A2          | B1          | B2         | C1         | Total       |
|---------------------|-----------|-------------|-------------|------------|------------|-------------|
| Root (orthographic) | 863 (863) | 1577 (1074) | 2247 (1110) | 2374 (866) | 1952 (528) | 9018 (4441) |
| Inflectional ending | 15 (15)   | 23 (9)      | 21 (2)      | 26 (4)     | 22 (3)     | 108 (33)    |
| Suffix              | 87 (87)   | 125 (57)    | 170 (61)    | 170 (39)   | 157 (29)   | 710 (273)   |
| Prefix              | 40 (40)   | 59 (33)     | 91 (37)     | 82 (28)    | 87 (23)    | 359 (161)   |
| Fogemorfem          | 4 (4)     | 5 (1)       | 6 (2)       | 7 (1)      | 6 (0)      | 28 (8)      |
| Unknown             | 12 (12)   | 30 (28)     | 55 (44)     | 36 (28)    | 25 (12)    | 158 (124)   |

Det här är baserat på typer i listan (inte förekomster).

26 jan. 2021

55

## Utveckling – morfologisk komplexitet / CEFR-nivå

### Morphological development at different levels

The first row in the table below is the number of words per level in LEGATO and the second row is the total number of annotated morphemes.

| Item                      | A1   | A2   | B1    | B2    | C1   | Total |
|---------------------------|------|------|-------|-------|------|-------|
| word                      | 1353 | 2646 | 4414  | 4316  | 3122 | 15854 |
| morpheme (Inte real root) | 2394 | 5559 | 10918 | 11253 | 8921 | 39053 |
| morpheme/word             | 1.77 | 2.10 | 2.47  | 2.61  | 2.86 | 2.46  |

Det här är alltså  
förekomster i listan.

26 jan. 2021

56

### Morpheme/CEFR level

Enter a morpheme or select one from the *drop-down* list to view the total number of occurrence of the selected morpheme in different L2 dataset.

Select Morpheme:

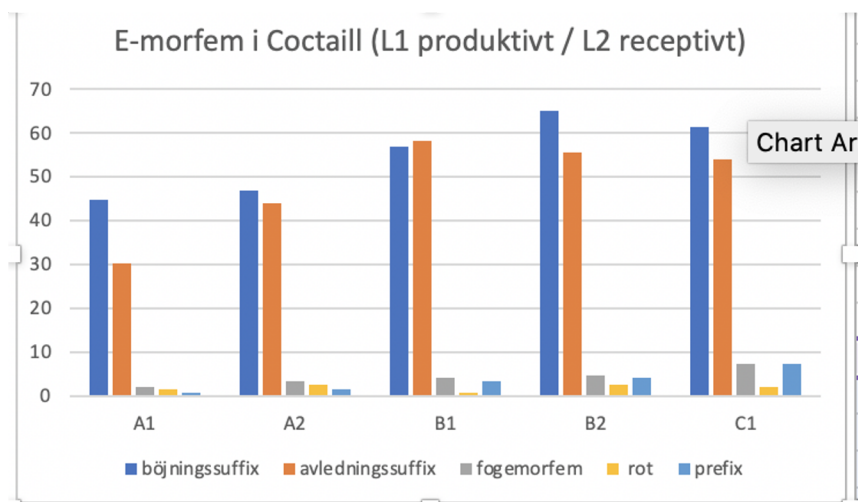
| Corpus      | A1  | A2  | B1  | B2  | C1  | C2 | Total |
|-------------|-----|-----|-----|-----|-----|----|-------|
| Legato      | 15  | 44  | 118 | 108 | 68  | 0  | 353   |
| Coctail     | 139 | 261 | 595 | 619 | 752 | 0  | 2366  |
| Swell-pilot | 27  | 82  | 56  | 86  | 97  | 7  | 355   |

[Download full set of comparison](#)

57

## E-morfem (per 10 000 token)

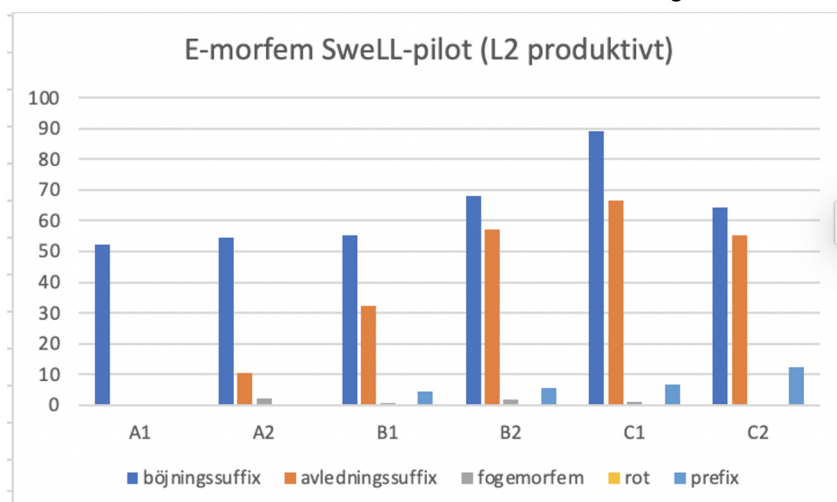
Böjning: pojke, pojkar  
 Avledning: rike, riken, riksdag  
 Rot: e-post  
 Fogemorfem: räknesätt, rikedom



58

## E-morfem (per 10 000 token)

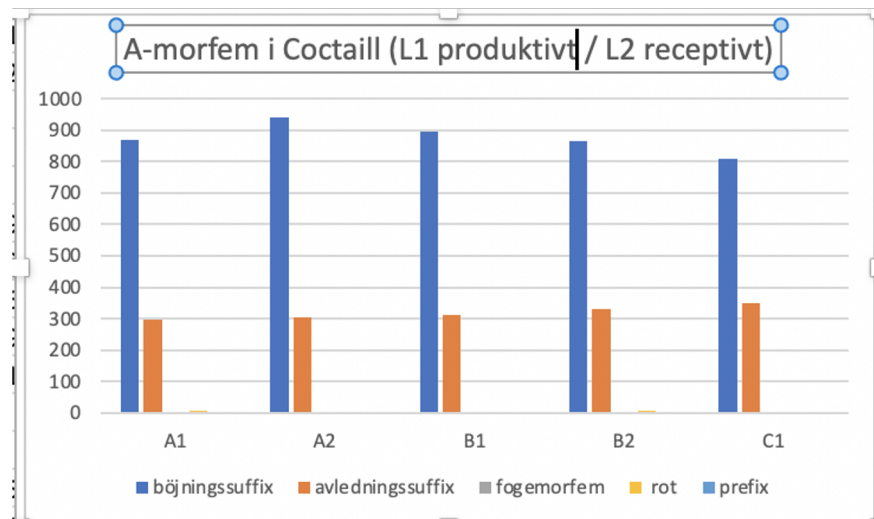
Böjning: pojke, pojkar  
 Avledning: rike, riken, riksdag  
 Rot: e-post  
 Fogemorfem: räknesätt, rikedom



59

## A-morfemen (per 10000 token)

Böjning: flicka, flickor; leva, lever  
 Avledning: Kenya, erinran; mata, matar  
 Fogemorfem: tålamod, vikingaskepp  
 Rot: a-människa, a  
 Prefix: amortering

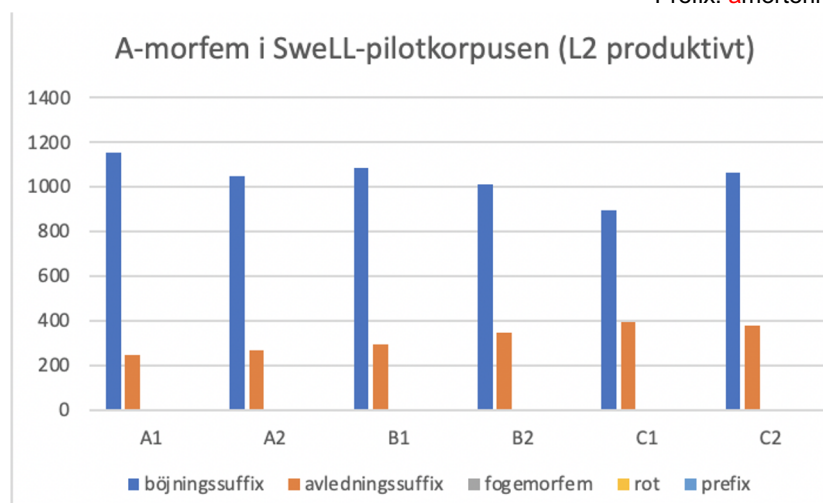


Baserat på orden som annoterats i Legato och deras frekvens i korpusen.

60

## A-morfemen (per 10000 token)

Böjning: flicka, flickor; leva, lever  
 Avledning: Kenya, erinran; mata, matar  
 Fogemorfem: tålamod, vikingaskepp  
 Rot: a-människa, a  
 Prefix: amortering



Baserat på orden som annoterats i Legato och deras frekvens i korpusen.

61

## Jfr English profile

- Utifrån våra 16 324 ord (lemgram+POS+Saldobetydelse) → när morfem som ingår förekommer
- Vilka ord de ingår i på vår lista
- Hur frekventa de är på olika nivåer (Coctail och SweLL-pilot) baserat på de här ordens frekvens i korpusarna

62

## skap = rot och avledningssuffix

Rot:

*nyskapa, skap, skapa, skapelse,  
ångestskapande, återskapa*

8 / 16324 annoterade ord

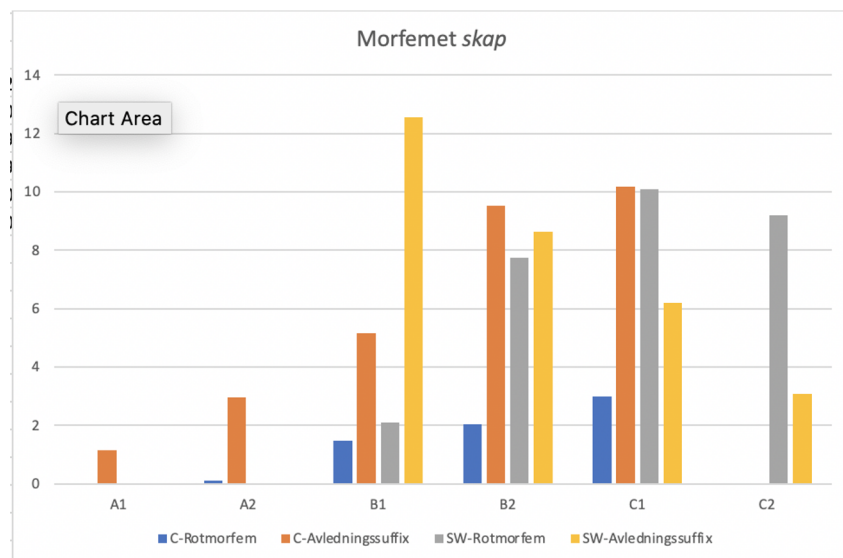
Avledningssuffix:

*kunskap, kunskapsmässig,  
kändisskap, landskap, ledarskap,  
litteraturvetenskap, medborgarskap,  
redskap*

53 / 16324 annoterade ord

63

## Hur vanligt är skap på de olika CEFR-nivåerna? (per 10000)



64

## I vilka ord förekommer skap på de olika nivåerna

Informationen finns i våra data – men ännu svåra att lätt sammanställa per morfem.

65

## I startgroparna...

- Hur utvecklar sig morfemen över CEFR-nivåerna?
- Kan vi liksom tidigare studier se en korrelation mellan morfemkännet och färdighet (receptiv och produktiv)?
- Kan det här användas för att bedöma nivå automatiskt?
- Kan ord automatiskt annoteras genom att tränas på våra annoteringar? (Yousuf jobbar på detta.)
- Sidoprojekt: Vilka morfem känner L2- (och L1-) talare faktiskt igen?
  - Undermedvetet vs medvetet
  - Ministudie på introkurs (L1 och L2 svenska)

66



P17-0716:1

L2Profiles

# Tack!

L2 profiler

Therese Lindström Tiedemann\*, Beatrice Silén\*, Stellan Petersson<sup>1</sup>, Maisa Lauriala\*

\*=Helsingfors universitet, <sup>1</sup> = Göteborgs universitet

Projektledare: Elena Volodina (Göteborgs universitet)

SPRÅKBANKEN **TEXT**



UNIVERSITY OF GOTHENBURG



UNIVERSITY OF HELSINKI

67

## Referenser

- Alfter, D., Tiedemann, T. L., & Volodina, E. (2019). LEGATO: A flexible lexicographic annotation tool. In *Proceedings of the 22nd Nordic Conference on Computational Linguistics* (pp. 382-388). <https://www.aclweb.org/anthology/W19-6145/>
- Amirjalili, F., & Jabbari, A. A. (2018). The impact of morphological instruction on morphological awareness and reading comprehension of EFL learners. *Cogent Education*, 5(1), 1523975.
- Bakken, K. 2002. Navnstatus og bestemdhetskategorien. NORNA:rapporter 29, särtryck av Namn och bygd.
- Balling, L. W. (2016). Findes morfemer? Om morfologiens rolle i sprogprocessering. *NyS, Nydanske Sprogstudier*, (51), 9-30.
- Bauer, L., & Nation, P. (1993). Word families. *International journal of Lexicography*, 6(4), 253-279.
- Brezina, V., & Pallotti, G. (2019). Morphological complexity in written L2 texts. *Second language research*, 35(1), 99-119.
- Brinton, L. J., & Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge University Press.
- Clahsen, H. (1985). 12 Profiling second language development: A procedure for assessing L2 proficiency. *Modelling and assessing second language acquisition*, 18, 283.
- Clahsen, H., Hansen, D., Ball, P., Crystal, M., & Fletcher, D. (2012). Profiling linguistic disability in German-speaking children. *Assessing grammar: The languages of LARSP*, 77-91.
- Crystal, D., Fletcher, P., & Garman, M. (1976). *The grammatical analysis of language disability: A procedure for assessment and remediation* (Vol. 1). Hodder Education.

68

- Granfeldt, J., & Ågren, M. (2014). SLA developmental stages and teachers' assessment of written French: Exploring Direkt Profil as a diagnostic assessment tool. *Language Testing*, 31(3), 285-305.
- Hopper, P. & Traugott, E (2003) *Grammaticalization*. 2:a utg. Cambridge university press.
- Laufer, B. (2005). Lexical frequency profiles: From Monte Carlo to the real world: A response to Meara (2005). *Applied linguistics*, 26(4), 582-588.
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied linguistics*, 16(3), 307-322.
- Lehtonen, M., Niska, H., Wande, E., Niemi, J., & Laine, M. (2006). Recognition of inflected words in a morphologically limited language: Frequency effects in monolinguals and bilinguals. *Journal of Psycholinguistic Research*, 35(2), 121.
- Löfdahl, M., Tingsell, S. och Wenner, L. 2015. Lexikon, onomastikon och flerspråkighet. I: E. Aldrin, L. Gustafsson, M. Löfdahl & L. Wenner (red.) *Innovationer i namn och namnmönster*. NORNA-rapporter 92 [https://www.isof.se/download/18.30cc95d915aacd94099dbf0/1529493974911/110%20art%20Lo%CC%88fdahl%20mfl%20\(002\).pdf](https://www.isof.se/download/18.30cc95d915aacd94099dbf0/1529493974911/110%20art%20Lo%CC%88fdahl%20mfl%20(002).pdf) Hämtad: 26 april 2020
- Löfdahl, M., Tingsell, S. och Wenner, L. 2016. Möten, motstånd och gränser - flerspråkiga namnstrategier. I: *Svenskans beskrivning* 35. S. 202-??
- Löfdahl, M. & Tingsell, S. 2013 [2016]. *Gröna gården och Häxeskogen*. Onomastikon och namnbildning ur ett andraspråksperspektiv. I. L. Mattfolk, M. Vidberg & P. Gustavsson (red.) *Namn i stadsmiljö*. NORNA-rapporter 90, Helsingfors. S. 169-187.

69



- Mochizuki, M., & Aizawa, K. (2000). An affix acquisition order for EFL learners: An exploratory study. *System*, 28(2), 291-304.
- Motschenbacher, H. 2020. Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names, *Names*, 68:2, 88-103, DOI:10.1080/00277738.2020.1731240
- Nagy, W. E., Carlisle, J. F., & Goodwin, A. P. (2014). Morphological knowledge and literacy acquisition. *Journal of learning disabilities*, 47(1), 3-12
- Nation, I. S. P. (1983) Testing and teaching vocabulary. *Guidelines*. 5: 12–25.
- Nyström, S. 1997. Namnmönster, namnsystem, onomastikon: Något om namnförrådets art och struktur. *Ellevte Nordiske Navneforskerkongressen. Sundvollen 19.-23. Juni 1994.*, s. 133-248. .
- Nyström, S. 2016. Names and meaning. I: C. Haugh (red.) *Oxford Handbook of Names and Naming*. Oxford: OUP. s. 40–51.
- Sasao, Y., & Webb, S. (2017). The word part levels test. *Language Teaching Research*, 21(1), 12-30.
- Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework: Word associations and verbal suffixes. *Studies in second language acquisition*, 19(1), 17-36.
- Sukying, A. (2020). Word Knowledge through Morphological Awareness in EFL Learners. *TESOL International Journal*, 15(1), 74-85.
- Svenska.se: SAOL, SO och SAOB
- Teleman, U., Hellberg, S., & Andersson, E. (1999). *Svenska akademiens grammatik*. Svenska akademien.
- Tyler, A., & Nagy, W. (1989). The acquisition of English derivational morphology. *Journal of memory and language*, 28(6), 649-667.
- Verhoeven, L., & Carlisle, J. F. (2006). Introduction to the special issue: Morphology in word identification and word spelling. *Reading and Writing*. 19(7), 643-650.

70

## Datakällor

Volodina, E., Pilán, I., Eide, S. R., & Heidarsson, H. (2014, November). You get what you annotate: a pedagogically annotated corpus of coursebooks for Swedish as a Second Language. In *Proceedings of the third workshop on NLP for computer-assisted language learning* (pp. 128-144).

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. *arXiv preprint arXiv:1604.06583*.

Svenska.se

SAG

71



P17-0716:1

L2Profiles

# Tack!

L2 profiler

Therese Lindström Tiedemann\*, Beatrice Silén\*, Stellan Petersson<sup>1</sup>, Maisa Lauriala\*

\*=Helsingfors universitet, <sup>1</sup> = Göteborgs universitet

Projektledare: Elena Volodina (Göteborgs universitet)

SPRÅKBANKENTEXT



UNIVERSITY OF GOTHENBURG



UNIVERSITY OF HELSINKI

72

73