Investigating morphological complexity by means of iCALL log data

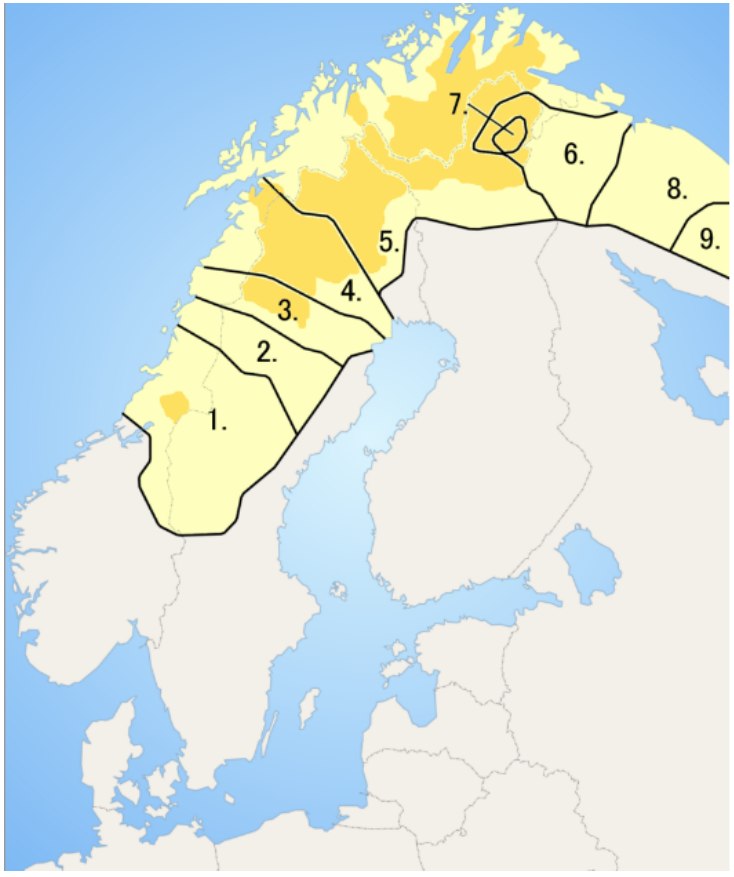
Trond Trosterud
Centre for Sami Language Technology
http://giellatekno.uit.no/

Contents

Introduction

- The idea is to treat log data from an interactive learning program as a corpus
  - N=300802 in the log, N=48252 for verbal morphology
- I have a twofold goal:
  - Investigate what causes trouble for the learners
  - Use the data to say something about morphological complexity in general.

The Sami Languages



- 1. South Sami
- 2. Ume Sami
- 3. Pite Sami
- 4. Lule Sami
- 5. North Sami
- 6. Skolt Sami
- 7. Inari Sami
- 8. Kildin Sami
- 9. Ter Sami

Darkened area represents municipalities that recognize Sami as an official language.

North Sami

- Morphologically complex – a suffixing language with many stem-changing processes
  - «a combination of Turkish and Icelandic»
- Inflects verbs in 3 persons and 3 numbers (sg, du, pl)
- A tense system similar to North and North-West Germanic
- 4 Moods: indicative, conditional, potential, (imperative)

North Sami vs. Finnish / Estonian

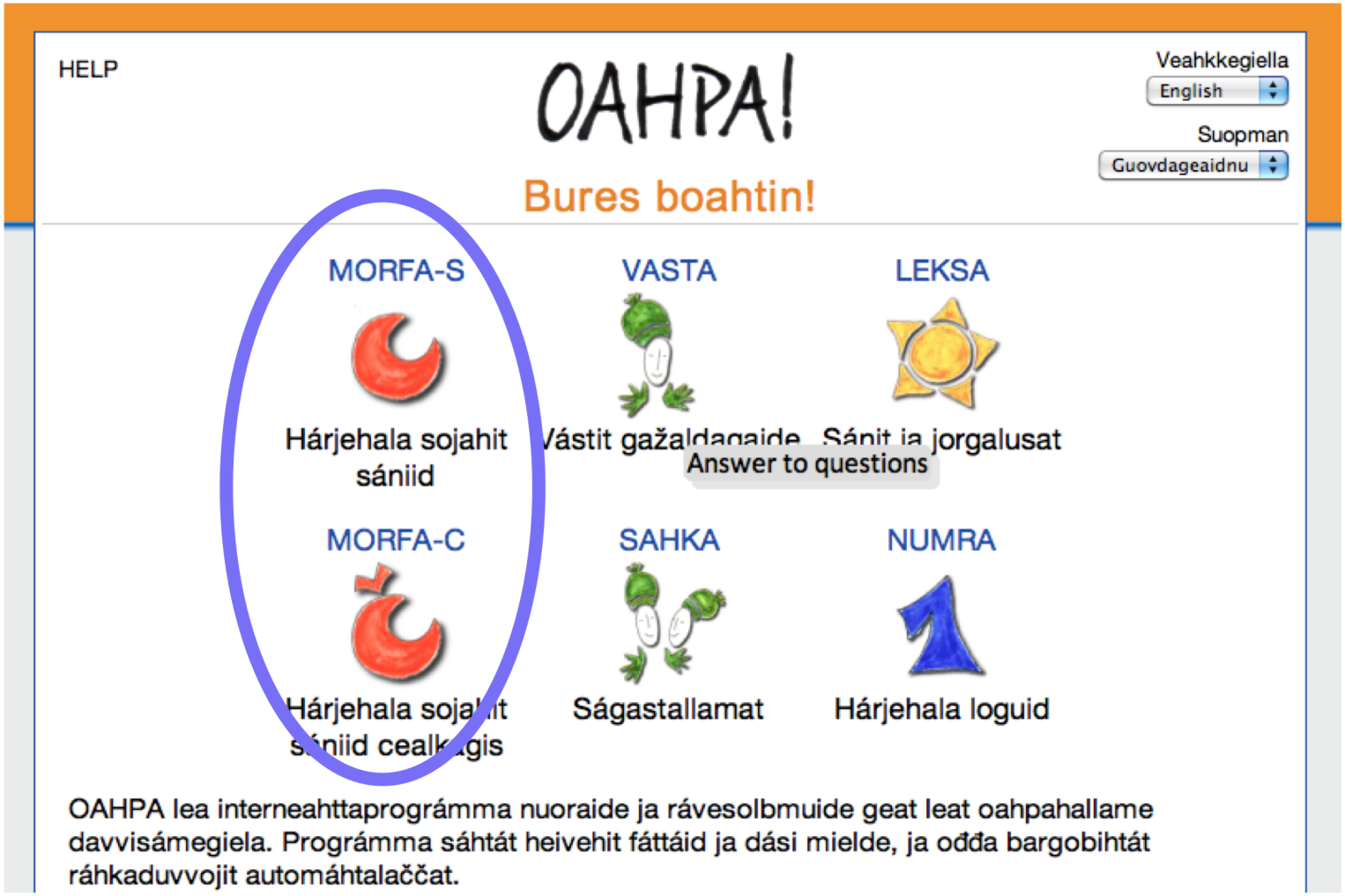
- Half the amount of cases
  - Two local cases: illative and a combined locative/elative
  - No partitive, no partial object distinction
- The verbal system quite similar
  - Same tense/negation system, but modes intertwined with tense
  - Fewer infinite forms

North Sami vs. Finnish / Estonian

- Half the amount of cases
  - Two local cases: illative and a combined locative/elative
  - No partitive, no partial object distinction
- The verbal system quite similar
  - Same tense/negation system, but modes intertwined with tense
  - Fewer infinite forms
- A more Estonian-like morphophonology
  - The genitive -n is lost
  - Consonant gradation includes almost all consonant groups
  - ... and shows a three-way quantity opposition

ICALL programs – http://oahpa.no/davvi/

ICALL programs – http://oahpa.no/



Morphology: Morfa

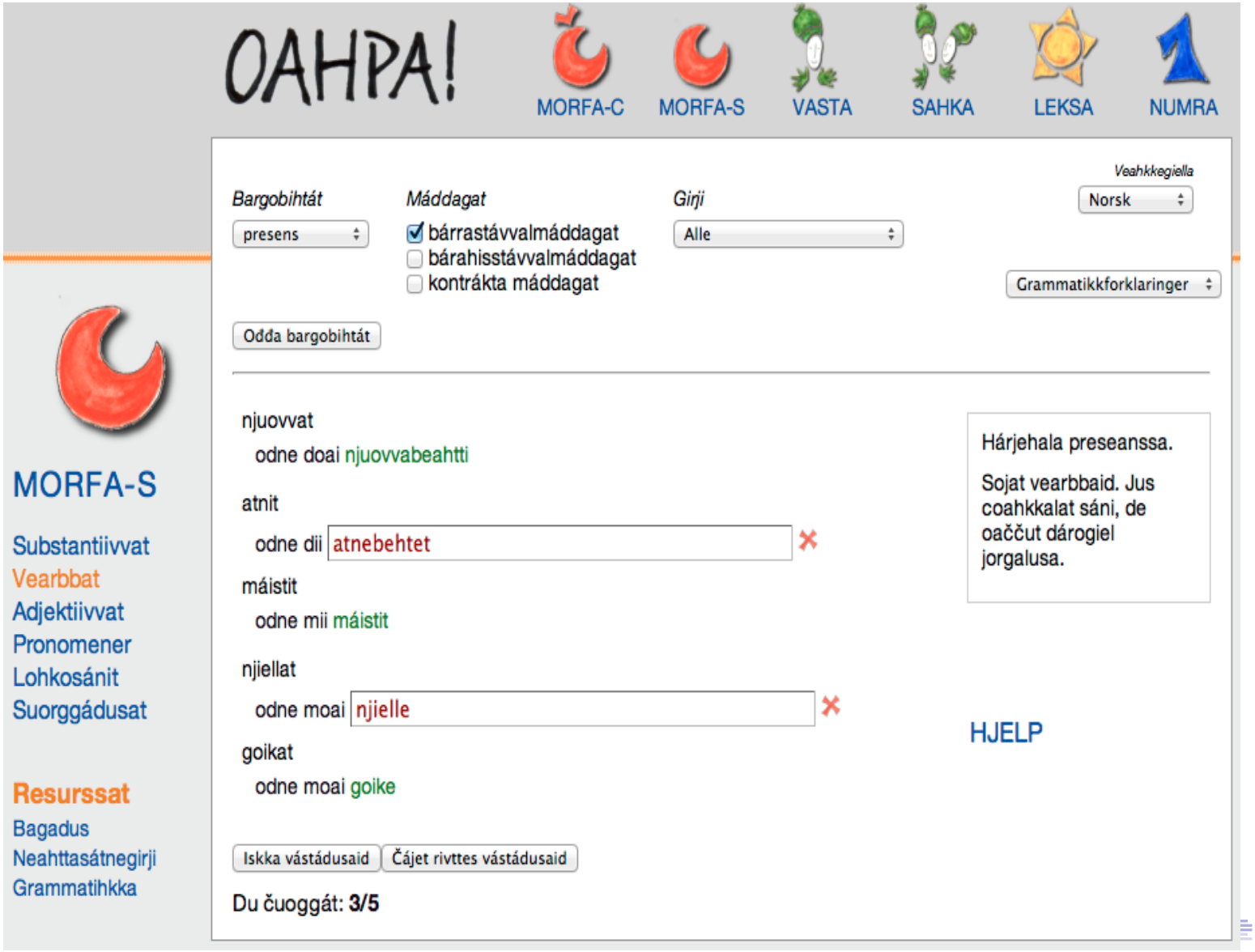
Morphology: Morfa

- Practice inflection
  - without context

The material

- 48252 verb queries, 29411 correct (60.95 %)
- 754 distinct verbs
- Tasks presented in sets x 5
- The task is to inflect the verb for person and number, based on pronoun

Morfa S


OAHPA!
MORFA-C
MORFA-S
VASTA
SAHKA
LEKSA
NUMRA
Bargobihtát
presens
Máddagat
bárrastávvalmáddagat
bárahisstávvalmáddagat
kontrákta máddagat
Girji
Alle
Veahkkegiella
Norsk
Grammatikkforklaringer
Ođđa bargobihtát
njuovvat
odne doai njuovvabeahtti
atnit
odne dii atnebehtet
máistit
odne mii máistit
njiellat
odne moai njielle
goikat
odne moai goike
Hárjehala preseanssa.
Sojat vearbbaid. Jus coahkkalat sáni, de oaččut dárogiel jorgalusa.
HJELP
MORFA-S
Substantiivvat
Vearbbat
Adjektiivvat
Pronomener
Lohkosánit
Suorggádusat
Resurssat
Bagadus
Neahttasátnegirji
Grammatihkka
Iskka vástádusaid
Čájet rivttes vástádusaid
Du čuoggát: 3/5

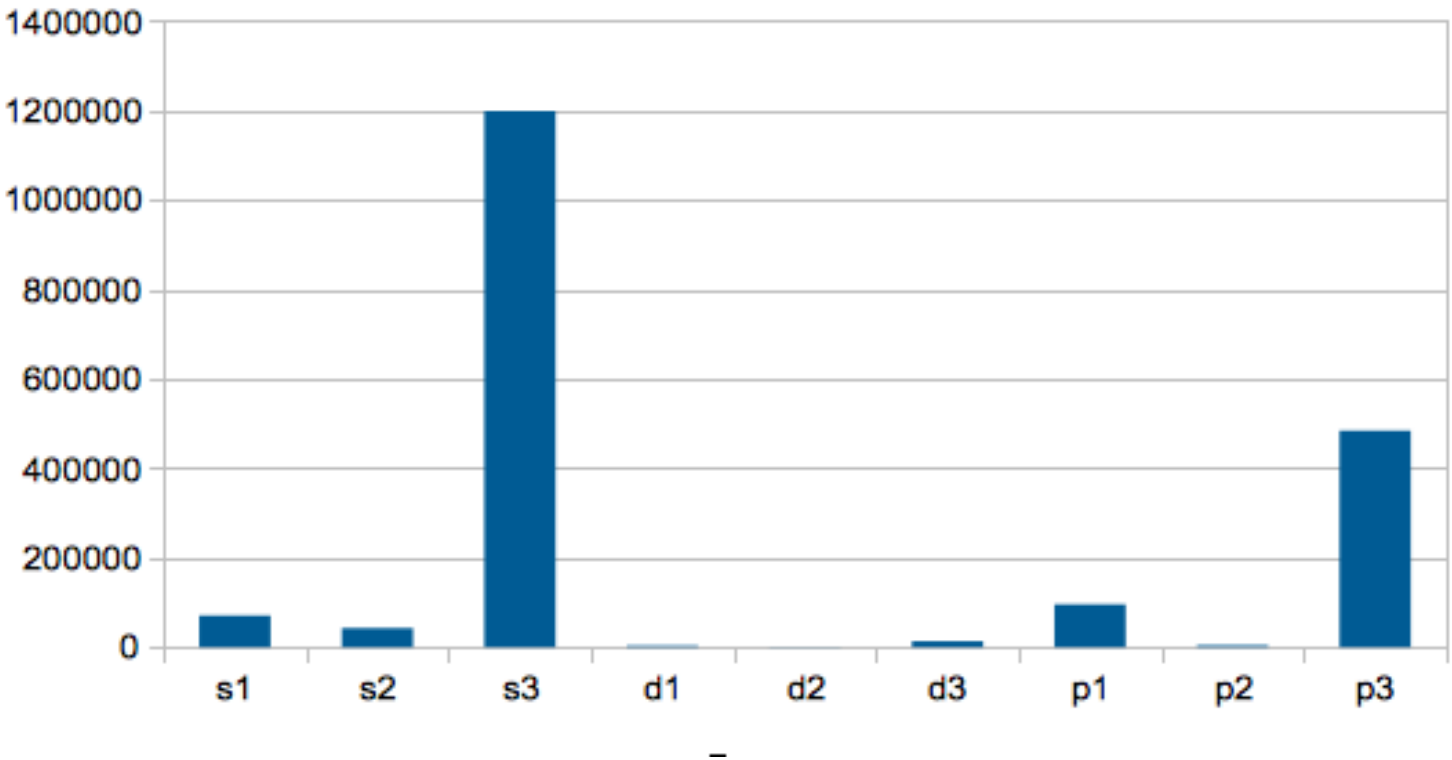
The log

```
1528|morfa_V|2009-01-26|dorret|0|doarrut|doarrut+V+Ind+Prs+Pl1|
1529|morfa_V|2009-01-26|doallaba|1|doallaba|doallat+V+Ind+Prs+Du3|
1530|morfa_V|2009-01-26|ruohttá|1|ruohttá|ruohttat+V+Ind+Prs+Sg3|
1531|morfa_V|2009-01-26|coggá|1|coggá|coggat+V+Ind+Prs+Sg3|
1532|morfa_V|2009-01-26|nohkat|1|nohkat|nohkkat+V+Ind+Prs+Sg2|
1533|morfa_V|2009-01-26|njuoskat|0|njuoskkat|njuoskat+V+Ind+Prs+Sg2|
1534|morfa_V|2009-01-26|deaivet|0|deivet|deaivat+V+Ind+Prs+Pl3|
1535|morfa_V|2009-01-26|borát|0|borat|borrat+V+Ind+Prs+Sg2|
1536|morfa_V|2009-01-26|deaivat|0|deivet|deaivat+V+Ind+Prs+Pl3|
1537|morfa_V|2009-01-26|borat|1|borat|borrat+V+Ind+Prs+Sg2|
1538|morfa_V|2009-01-26|oasta|0|oastá|oastit+V+Ind+Prs+Sg3|
1539|morfa_V|2009-01-26|jorribeahtti|0|jorrabeahtti|jorrat+V+Ind+Prs+Du2|
1540|morfa_V|2009-01-26|duostaba|1|duostaba|duostat+V+Ind+Prs+Du3|
```

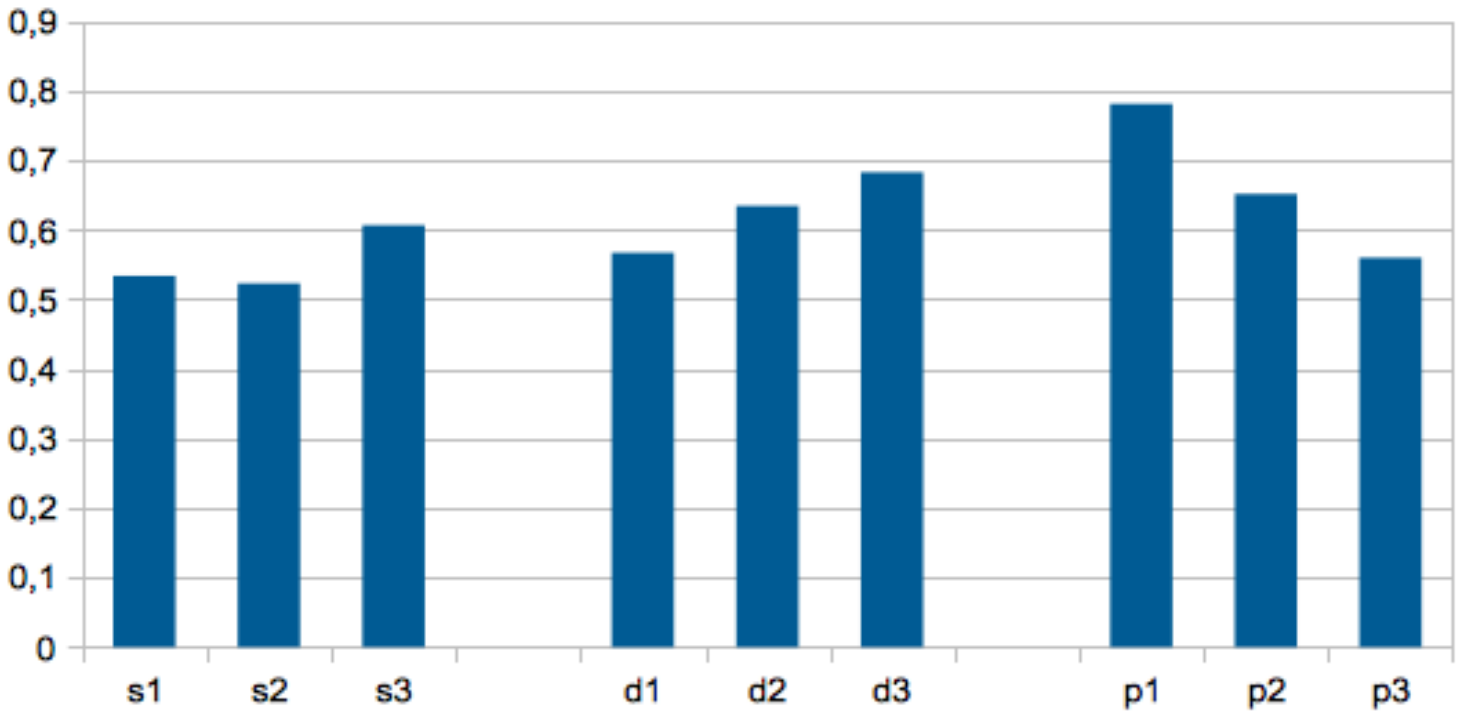
Some possible factors determining error patterns

1. Learning via immersion: rare forms are hard, common ones easier
2. Markedness: Marked morphological categories are harder
3. Some morphological processes are harder to perform than others

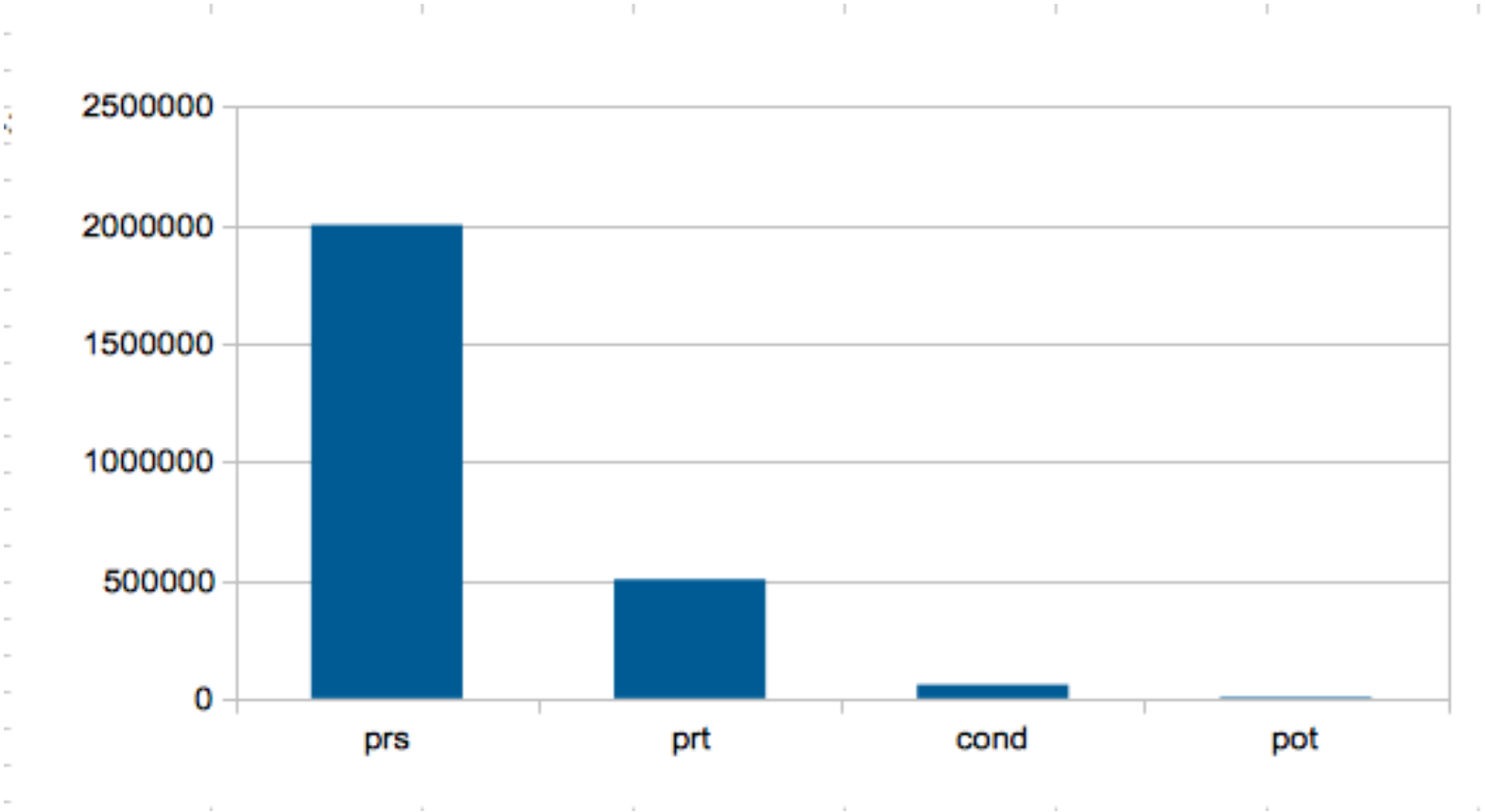
Usage in 20m corpus



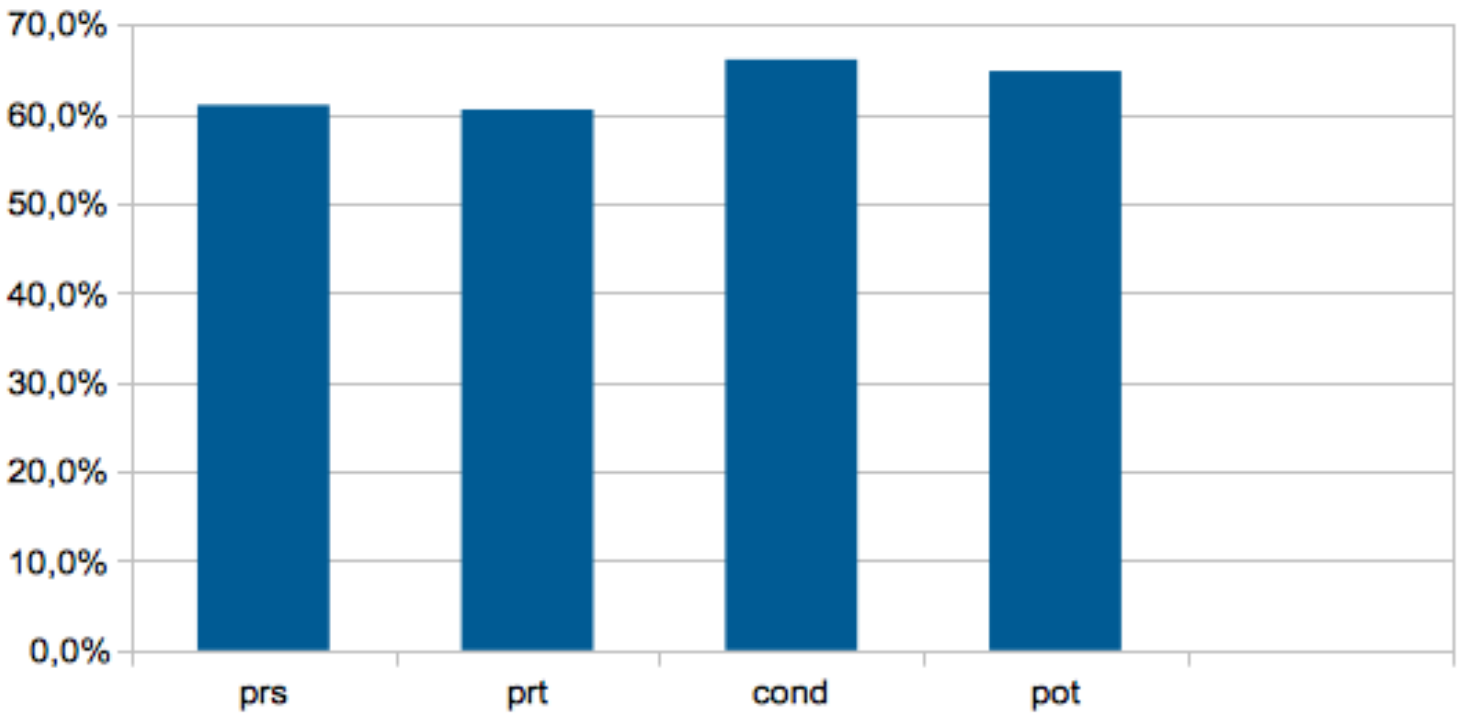
Correct answer in Oahpa



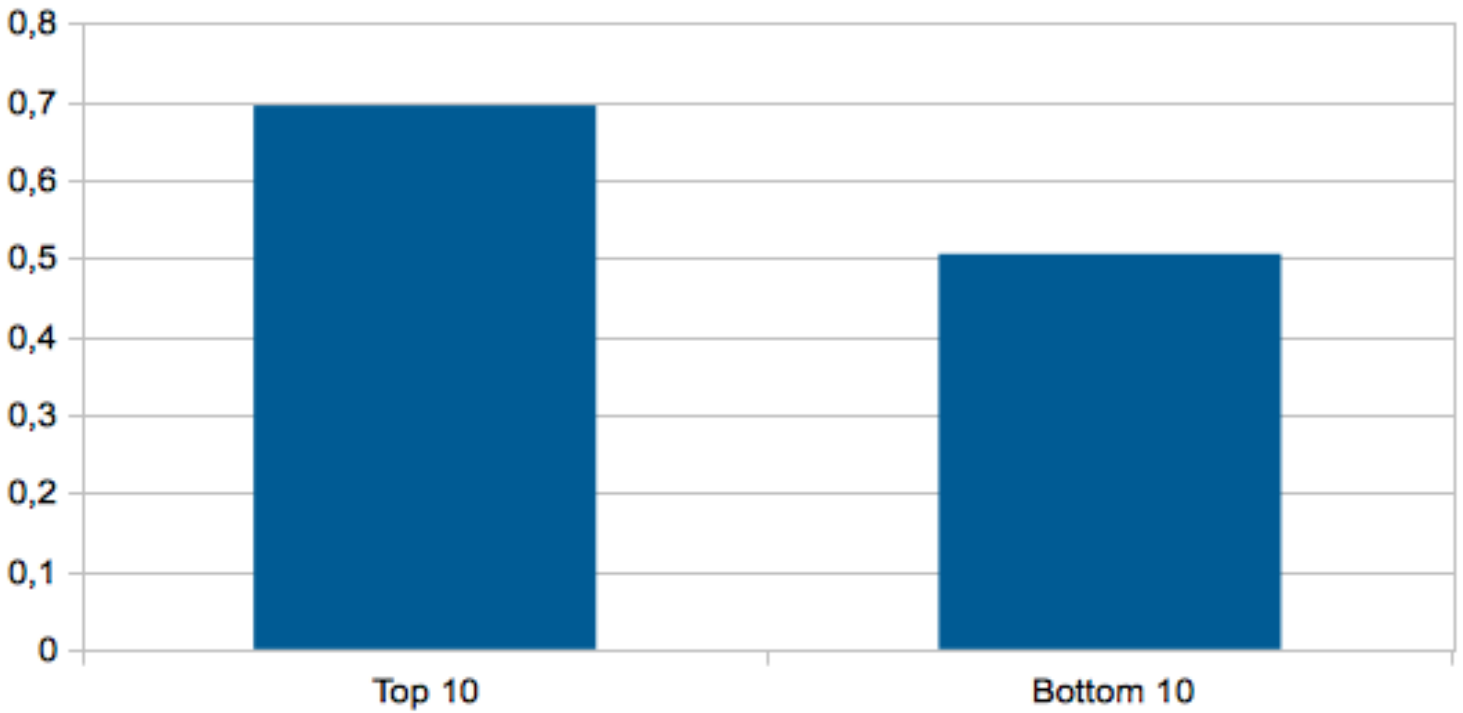
Usage in 20m corpus



Correct answer in Oahpa



Lexical frequency plays a rôle, though

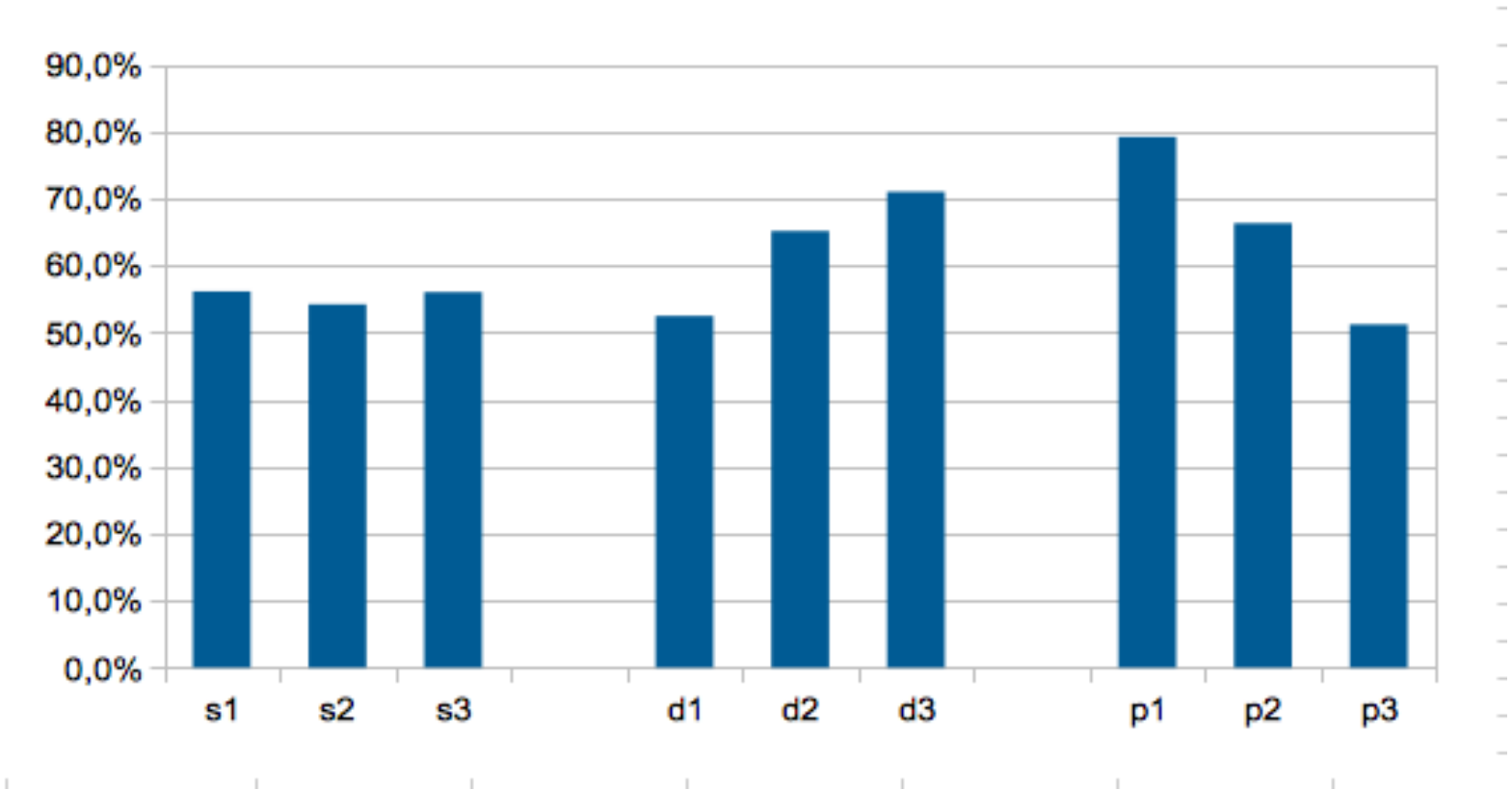


The learner's task

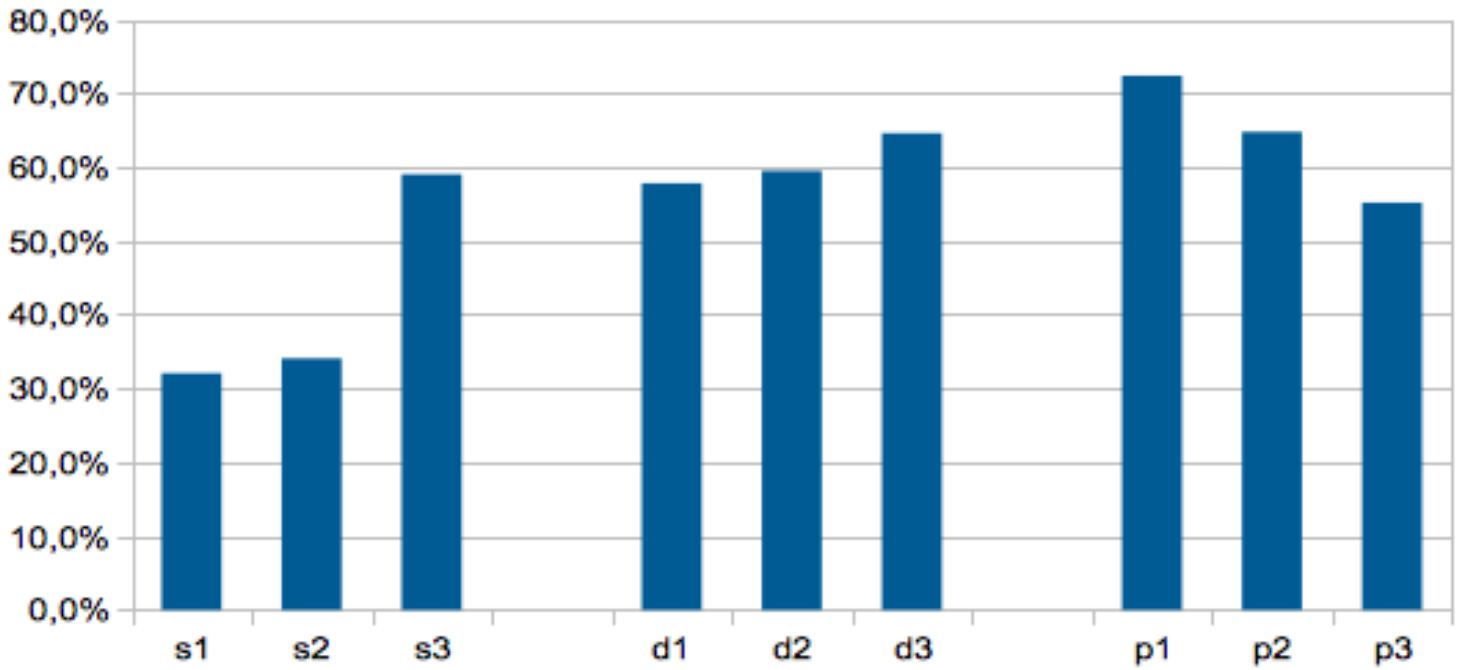
1. Add the correct suffix
2. Make the relevant consonant gradation change
3. If needed, change the root diphthong
4. If needed, change the stem final vowel

VOWEL STEM VERBS - present Shaded: weak grade			
Person	*viehkat* 'to run'	*diehtit* 'to know'	*goarrut* 'to sew'
mun	*viegan*	*dieđán*	*goarun*
don	*viegat*	*dieđát*	*goarut*
son	*viehká*	*diehtá*	*goarru*
moai	*vihke*	*dihte*	*gorro*
doai	*viehkabeahtti*	*diehtibeahtti*	*goarrubeahtti*
soai	*viehkaba*	*diehtiba*	*goarruba*
mii	*viehkat*	*diehtit*	*goarrut*
dii	*viehkabehtet*	*diehtibehtet*	*goarrubehtet*
sii	*vihket*	*dihtet*	*gorrot*

Bisyllabic -at/-ut present

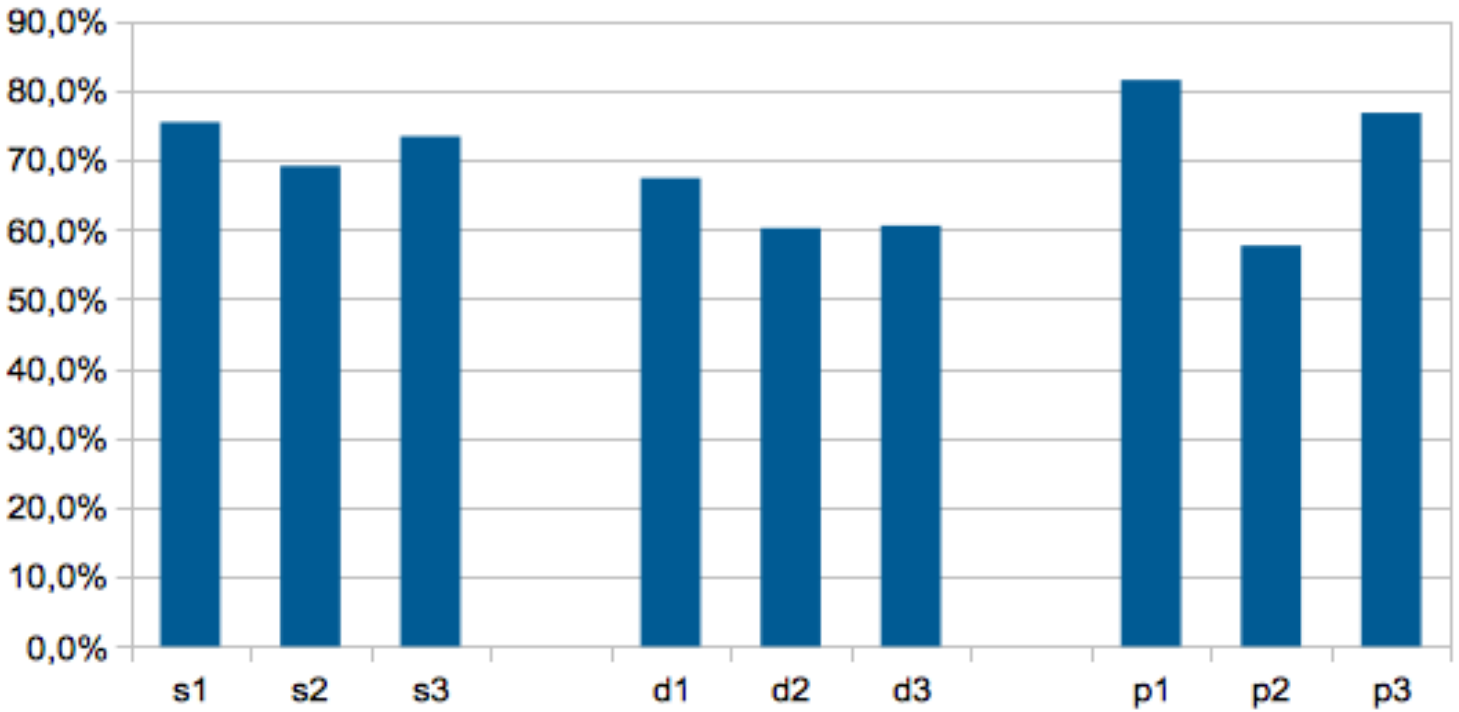


Bisyllabic -it present



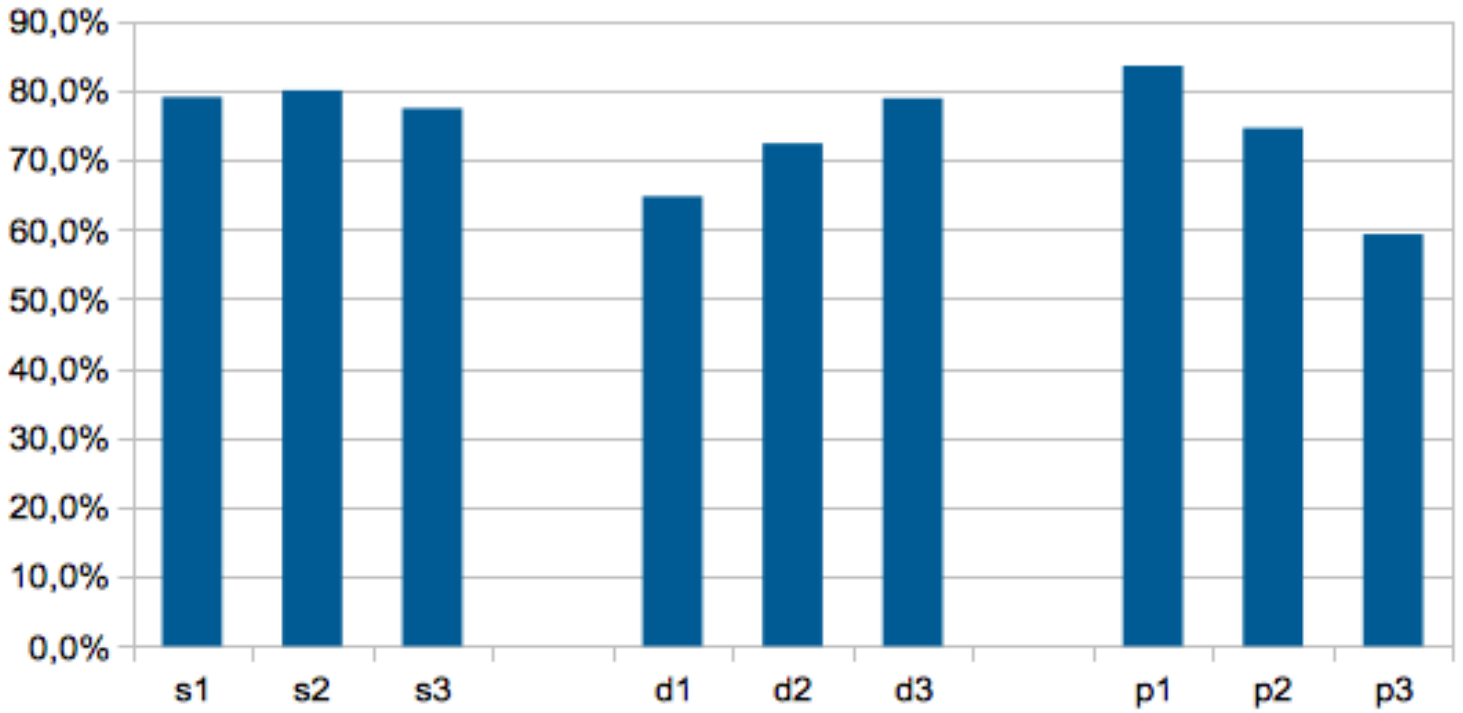
CONSONANT STEM VERBS - present		
Person	*muitalit* 'to tell'	*leat* 'to be'
mun	*muitalan*	*lean*
don	*muitalat*	*leat*
son	*muitala*	*lea*
moai	*muitaletne*	*letne*
doai	*muitaleahppi*	*leahppi*
soai	*muitaleaba*	*leaba*
mii	*muitalit, muitalat*	*leat*
dii	*muitalehpet*	*lehpet*
sii	*muitalit*	*leat*

Trisyllabic present



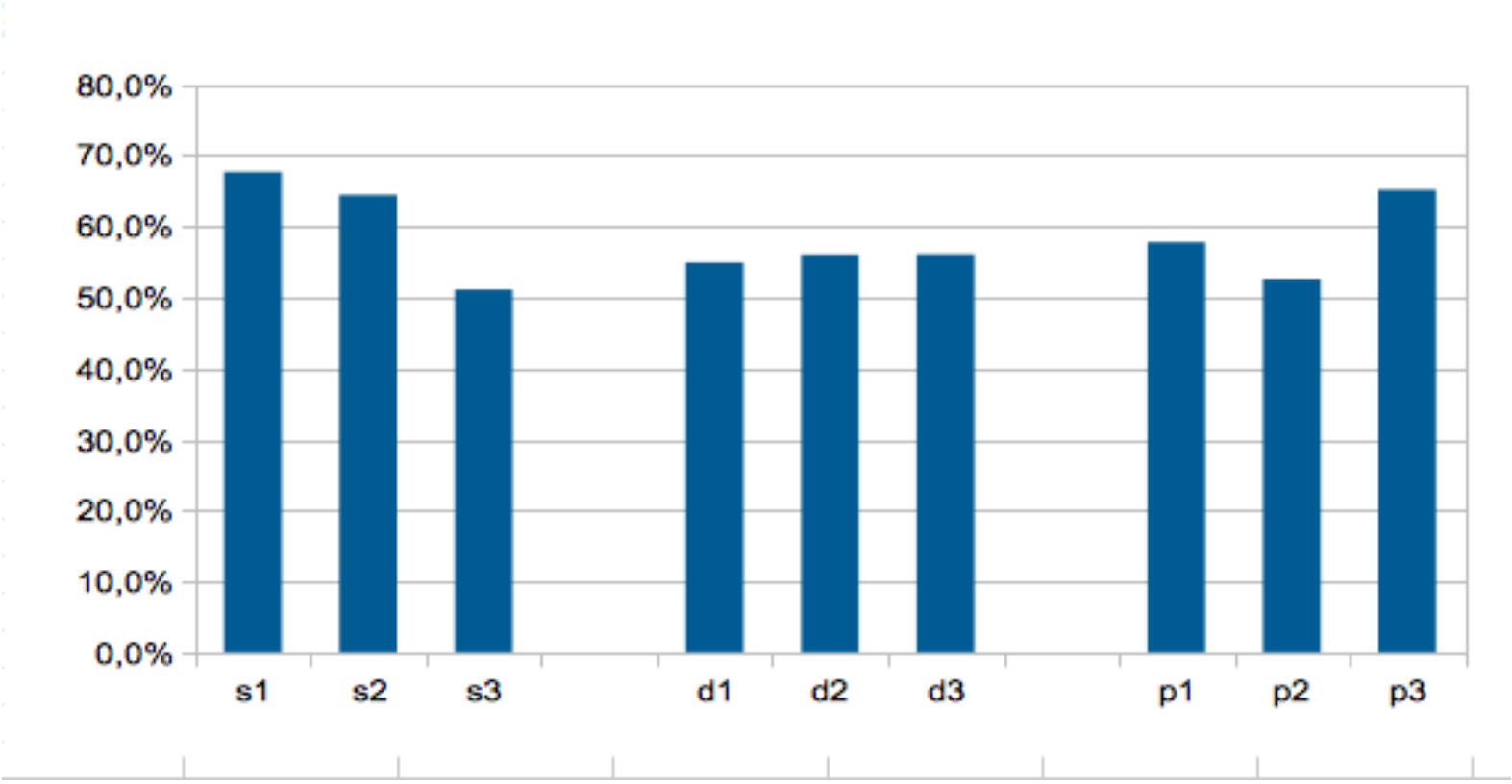
CONTRACTION VERBS - present			
Person	*čohkkát* 'to sit'	*fárret* 'to move'	*diŋgot* 'to order'
mun	*čohkkán*	*fárren*	*diŋgon*
don	*čohkkát*	*fárret*	*diŋgot*
son	*čohkká*	*fárre*	*diŋgo*
moai	*čohkkájetne*	*fárrejetne*	*diŋgojetne*
doai	*čohkkábeahtti*	*fárrebeahtti*	*diŋgobeahtti*
soai	*čohkkába*	*fárreba*	*diŋgoba*
mii	*čohkkát*	*fárret*	*diŋgot*
dii	*čohkkábehtet*	*fárrebehtet*	*diŋgobehtet*
sii	*čohkkájit*	*fárrejit*	*diŋgojit*

Contracted present

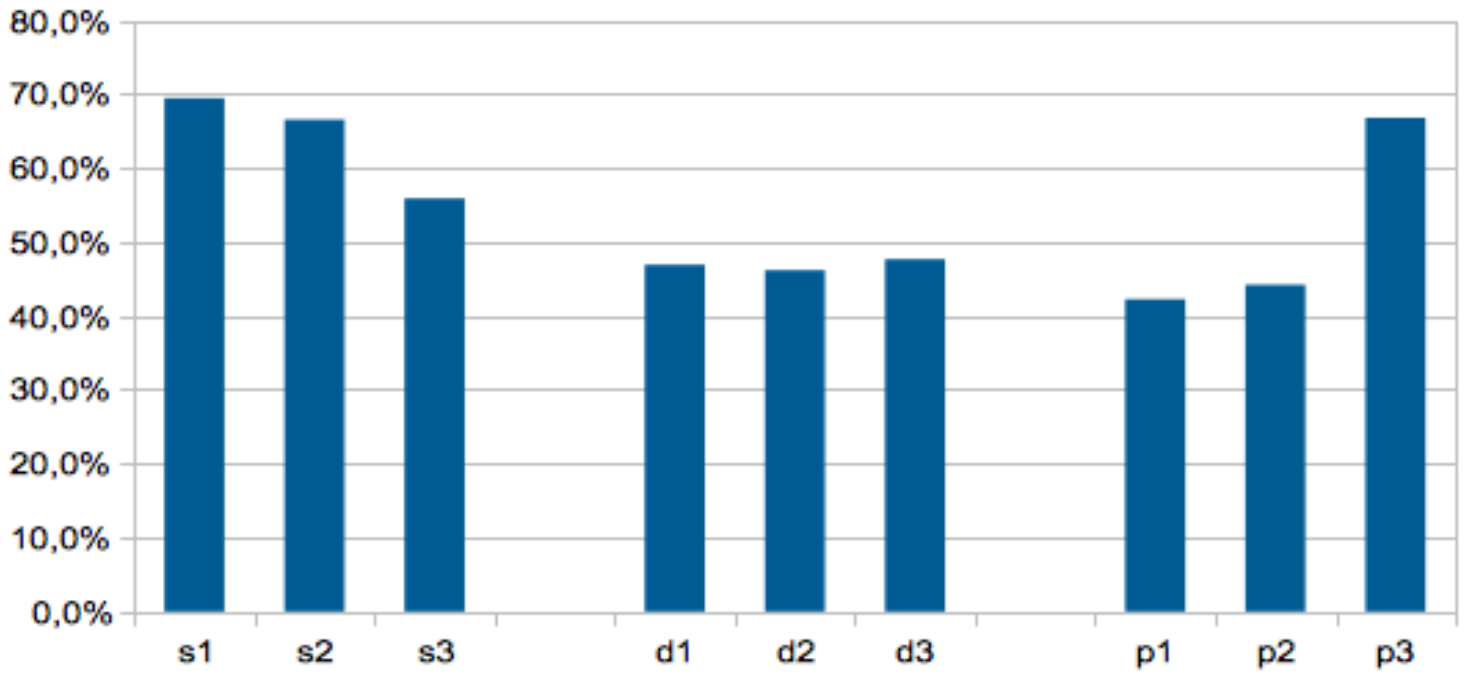


VOWEL STEM VERBS - preterite Shaded: weak grade			
Person	*viehkat* 'to run'	*diehtit* 'to know'	*goarrut* 'to sew'
mun	*vihken*	*dihten*	*gorron*
don	*vihket*	*dihtet*	*gorrot*
son	*viegai*	*diđii*	*goarui*
moai	*viegaime*	*diđiime*	*goaruime*
doai	*viegaide*	*diđiide*	*goaruide*
soai	*viegaiga*	*diđiiga*	*goaruiga*
mii	*viegaimet*	*diđiimet*	*goaruimet*
dii	*viegaidet*	*diđiidet*	*goaruidet*
sii	*vihke*	*dihte*	*gorro*

Bisyllabic -at/-ut preterite

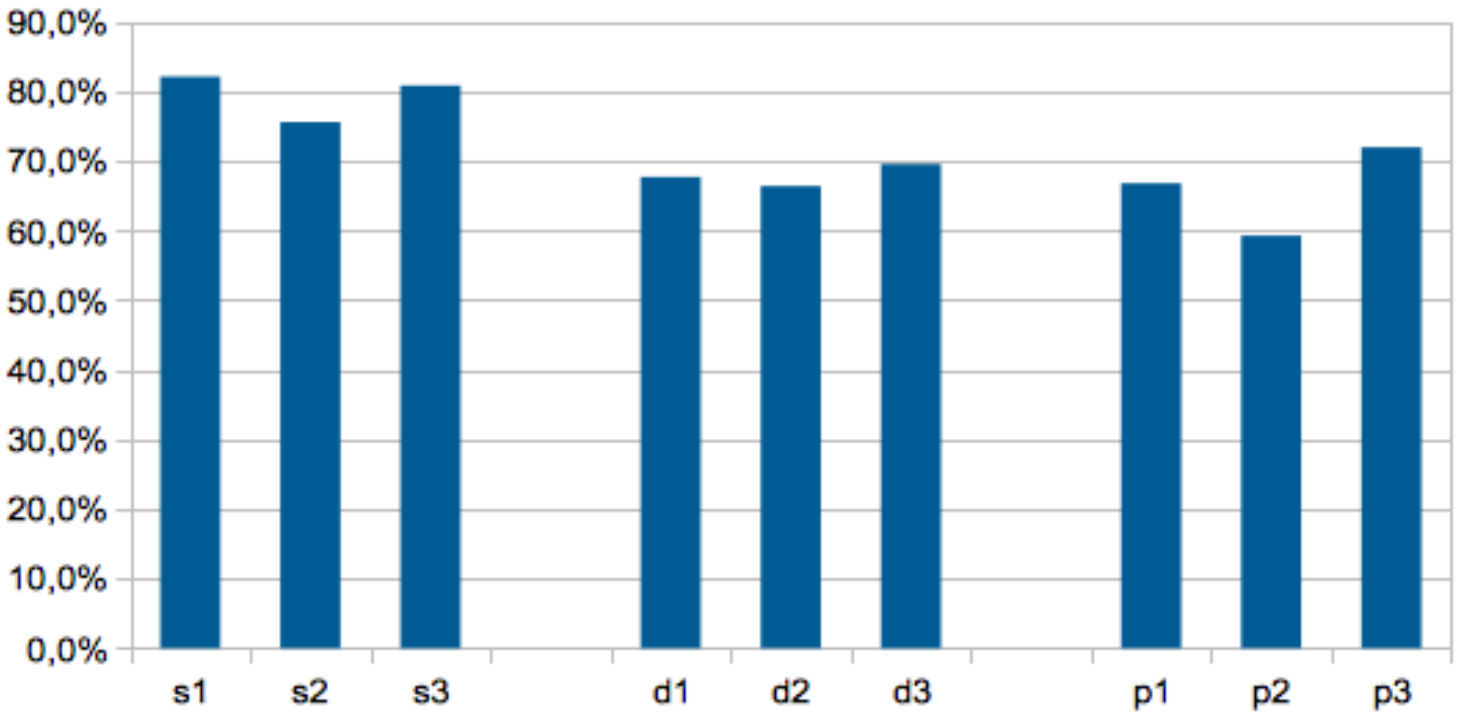


Bisyllabic -it preterite



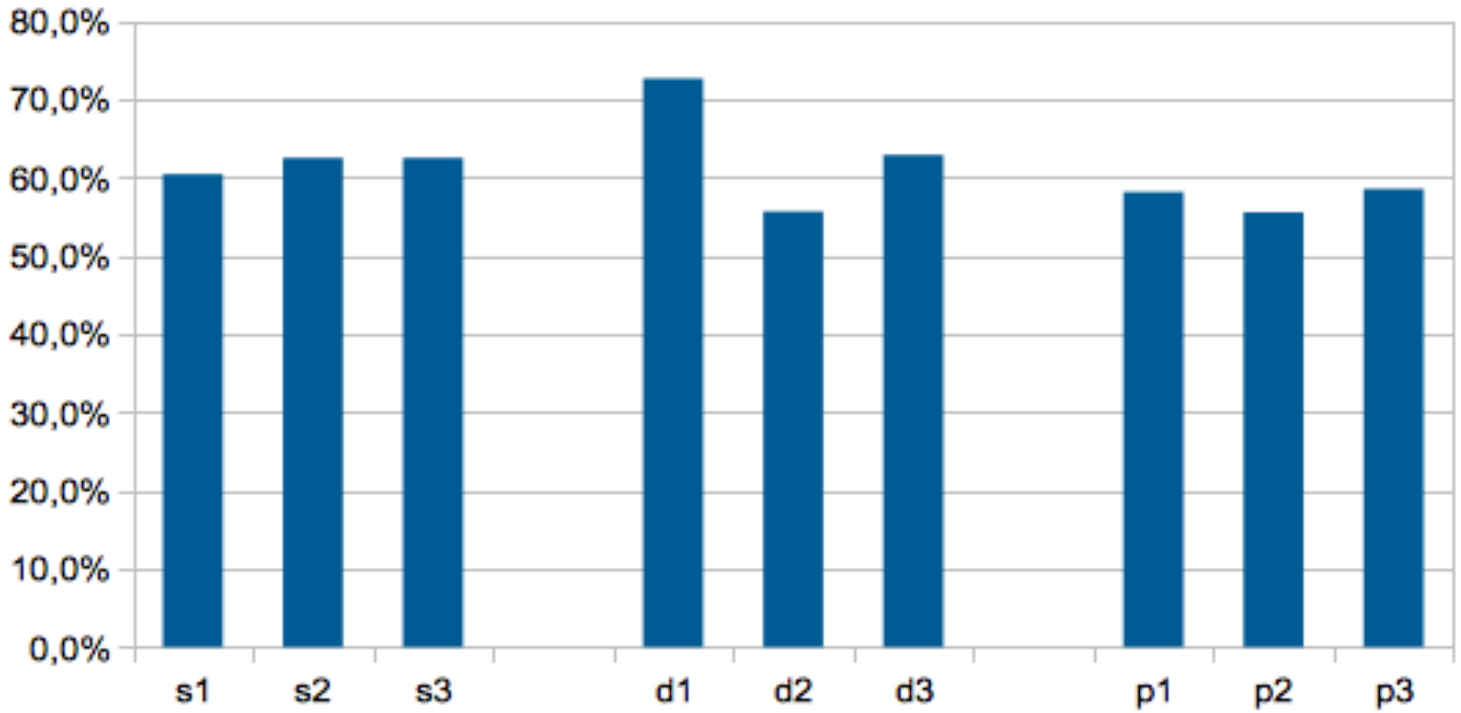
CONSONANT STEM VERBS - preterite		
Person	muitalit	leat
mun	muitalin	ledjen
don	muitalit	ledjet
son	muitalii	lei, leai
moai	muitaleimme	leimme
doai	muitaleidde	leidde
soai	muitaleigga	leigga
mii	muitaleimmet	leimmet
dii	muitaleiddet	leiddet
sii	muitaledje	ledje

Trisyllabic preterite



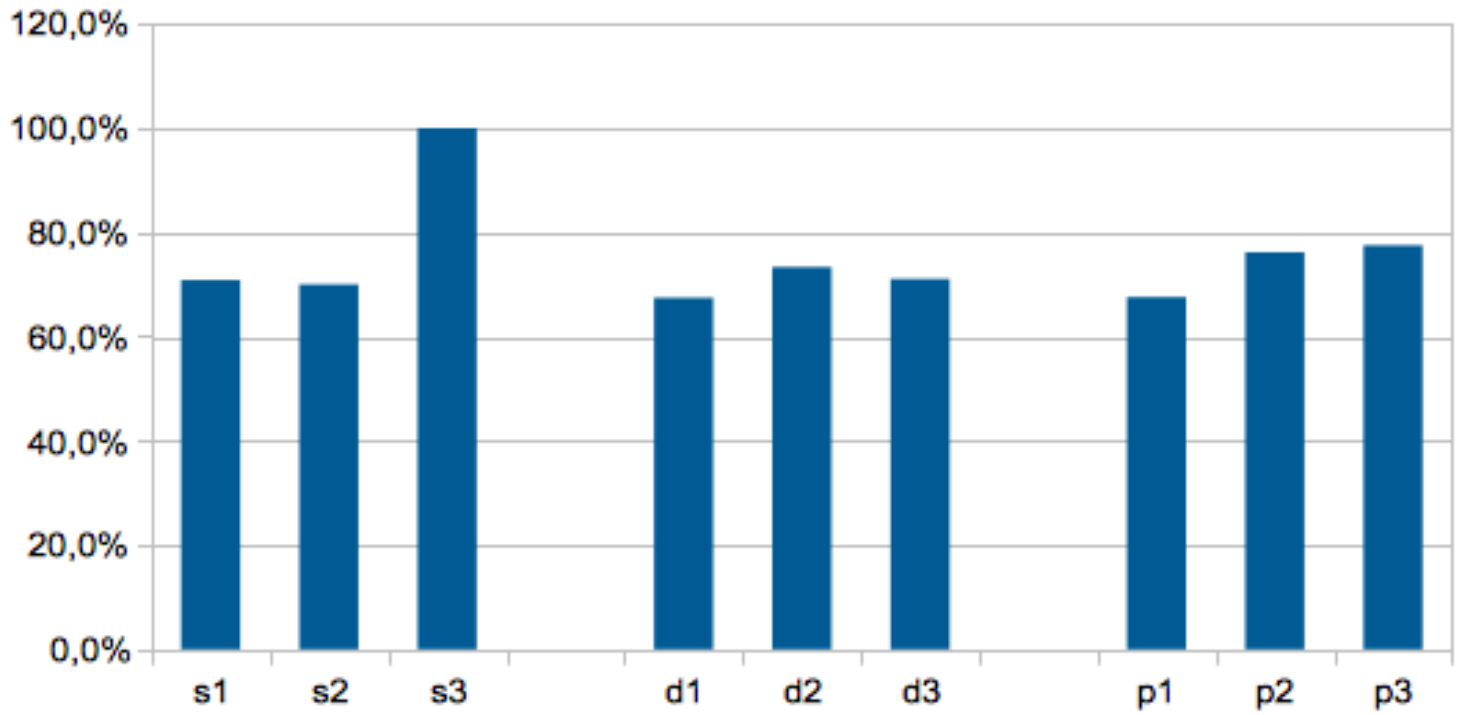
VOWEL STEM VERBS - conditional			
Person	*viehkat* 'to run'	*diehtit* 'to know'	*goarrut* 'to sew'
mun	*viegašin*	*dieđášin*	*gorošin*
don	*viegašit*	*dieđášit*	*gorošit*
son	*viegašii*	*dieđášii*	*gorošii*
moai	*viegašeimme*	*dieđášeimme*	*gorošeimme*
doai	*viegašeidde*	*dieđášeidde*	*gorošeidde*
soai	*viegašeigga*	*dieđášeigga*	*gorošeigga*
mii	*viegašeimmet*	*dieđášeimmet*	*gorošeimmet*
dii	*viegašeiddet*	*dieđášeiddet*	*gorošeiddet*
sii	*viegašedje*	*dieđášedje*	*gorošedje*

Bisyllabic conditional



CONSONANT STEM VERBS - conditional		
Person	muitalit	leat
mun	muitalivččen	livččen
don	muitalivččet	livččet
son	muitalivččii	livččii
moai	muitalivččiime	livččiime
doai	muitalivččiide	livččiide
soai	muitalivččiiga	livččiiga
mii	muitalivččiimet	livččiimet
dii	muitalivččiidet	livččiidet
sii	muitalivčče	livčče

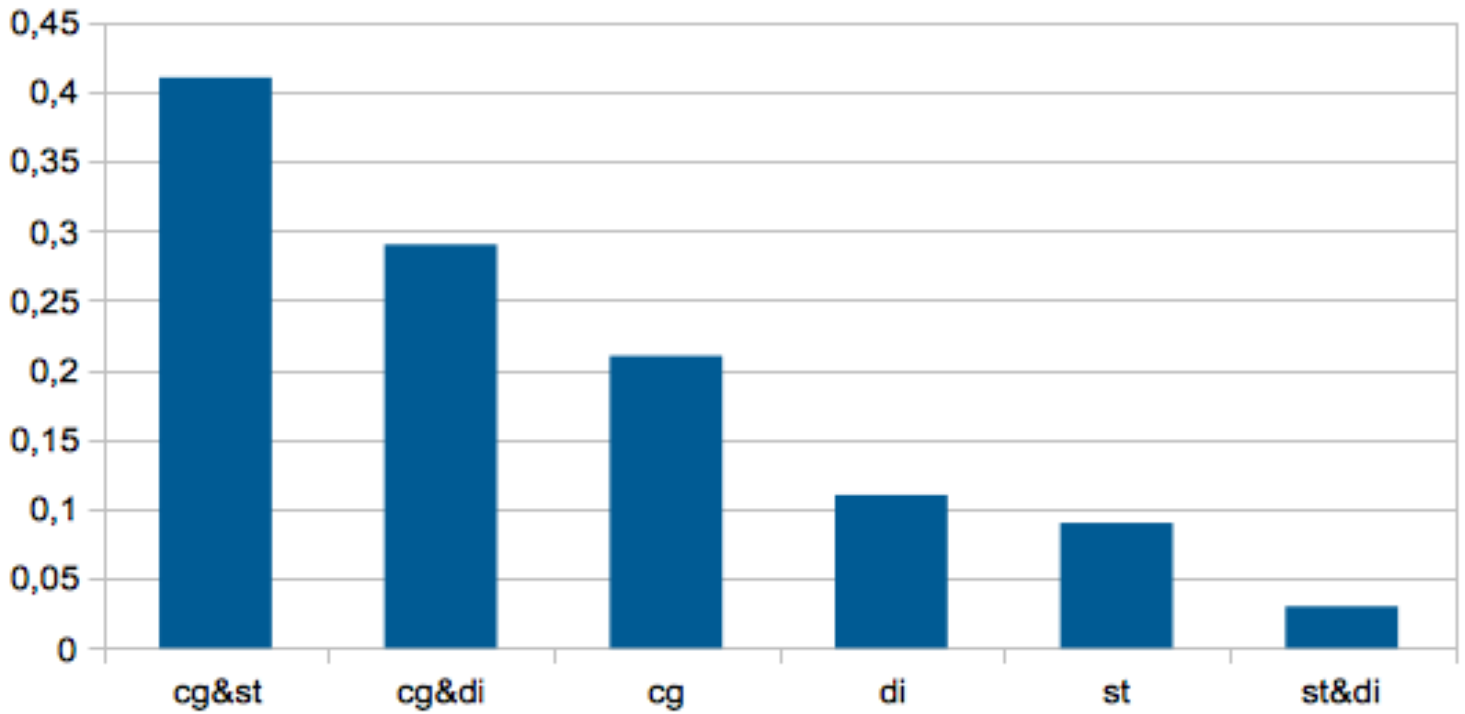
Trisyllabic conditional



Example cell: 1st person Present indicative of bisyllabic -at-verbs (N=1570)

- The correct form is *viehkat -> viegan*
- The log has 56% correct answers
- Generalisations from the errors:
  - 9.2% of the errors show wrong suffix (no *-n*)
  - 15.6% of the errors have hypercorrect *-án* (from -it verbs)
  - 8.9% of the errors have some other vowel before *-n*
  - 36.7% of the errors show strong grade for correct weak grade

The effect of the different morphological processes



Conclusion

Conclusion

- Relevance to learnability
  - Usage frequency is not relevant for predicting the difficulty of MS features per se
  - Lexeme frequency seems to have an effect

Conclusion

- Relevance to learnability
  - Usage frequency is not relevant for predicting the difficulty of MS features per se
  - Lexeme frequency seems to have an effect
- Relevance to morphological theory in general
  - Markedness hierarchy of morphological processes:
    - *Consonant gradation & other processes* > *Consonant gradation alone* > *Root vowel alternation* > *Stem vowel alternation* > *Suffixation* > *Conversion*

Conclusion

- Relevance to learnability
  - Usage frequency is not relevant for predicting the difficulty of MS features per se
  - Lexeme frequency seems to have an effect
- Relevance to morphological theory in general
  - Markedness hierarchy of morphological processes:
    - *Consonant gradation & other processes* > *Consonant gradation alone* > *Root vowel alternation* > *Stem vowel alternation* > *Suffixation* > *Conversion*
- Plans for future work
  - Conduct a statistical analysis on the data (*no friends among statisticians yet...*)
    - binomial regression analysis
    - look at effects of each verb in isolation
  - Link the data of future logs to individual students

Conclusion

- Relevance to learnability
  - Usage frequency is not relevant for predicting the difficulty of MS features per se
  - Lexeme frequency seems to have an effect
- Relevance to morphological theory in general
  - Markedness hierarchy of morphological processes:
    - *Consonant gradation & other processes* > *Consonant gradation alone* > *Root vowel alternation* > *Stem vowel alternation* > *Suffixation* > *Conversion*
- Plans for future work
  - Conduct a statistical analysis on the data (*no friends among statisticians yet...*)
    - binomial regression analysis
    - look at effects of each verb in isolation
  - Link the data of future logs to individual students

Kiitos !