Morphological variation in Swiss German: Two case studies

Manuela Schönenberger[1], Eric Haeberli[1] & Richard Zimmermann[2]

([1]University of Geneva, [2]University of Manchester)

UNIVERSITÉ DE GENÈVE
FACULTÉ DES LETTRES
Département de linguistique

FNSNF
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

MANCHESTER 1824
The University of Manchester

Soziolinguistische Perspektiven: Kontakt – Variation – Wandel

12.05.–13.05.23, Universität Bern

What the talk is about

- Focus on two phenomena of morphological variation in a variety of Swiss German (CH)

- Characteristics of the phenomena studied
  - Variables with two variants that do not seem to differ in meaning.
  - The type of variation is not found in German.

- The role different linguistic and non-linguistic factors play in the two areas of variation will be examined.

- In particular: Is there evidence for change in progress?

- Tentative conclusions

Background

- While diatopic variation in CH has been investigated extensively for various linguistic domains, much less is known about inter- and intra-speaker variation within specific varieties of CH, in particular as far as morphology and syntax are concerned.

- CH is well suited for the study of morphological/syntactic variation: A non-standard variety used naturally in everyday life by members of all social backgrounds; no normative pressures; limited influence of the standard.

- General aim: To gain a better understanding of morphological and syntactic variation by investigating inter- and intra-speaker variation in a specific variety of CH.

Two phenomena of morphological variation

- Indefinite neuter article:

  es (Chind) or *e (Chind)* 'a child'
  (German: no variation: *ein (Kind)*)

- 1SG present tense of *go* 'go':

  (i) gang or *(i) gò* 'I go'
  (German: no variation: *ich geh(e)*)

The data

- The data were retrieved from a corpus of spoken Swiss German: *WilKo – Geparstes Korpus von Spontansprachdaten des Schweizerdeutschen der Stadt Wil*.

- Production data from 62 speakers speaking the same CH dialect
  – local dialect of Wil (SG), 24 000 inhabitants
  – all data obtained from informal interviews of 90+ MIN
    (120h of audio-recordings, 1.4 million words)

- Choice of speakers
  – must be a native speaker of the local dialect spoken in Wil
    (growing up and attending school in Wil)
  – can be classified into one of the age groups, different social backgrounds

- Age groups
  young: aged 20–30   (18 speakers: 10f/8m) (average age: 24)
  middle-aged: aged 45–60   (20 speakers: 9f/11m) (average age: 54)
                 + 2 female interviewers
  elderly: aged 70+   (17 speakers: 8f/9m) (average age: 77)
  no group: (5 speakers: 4f (average age: 64) and 1m (aged 40–45))

Research question

- With both phenomena, the two variants seem to be semantically equivalent. What linguistic or non-linguistic factors determine the use of the two variants in each case?

Retrieval of data

- The audio data have been transcribed using EXMaRALDA (Schmidt & Wörner 2009).

- Since the parsing of the full corpus will only be completed later this year, the relevant data were retrieved with the tool "EXACT search" in EXMaRALDA, which uses grep (standard Unix pattern matching). The results of these searches were pasted into an Excel document, and then classified by hand according to various criteria.

Statistical methodology

- The data distribution of both case studies was modeled with a number of logistic regression models predicting the proportion of one form from a number of factors.

- Many models are possible. However, using trial and error, model evaluation metrics and hypothesis-guided common sense, one "best" model is proposed to capture the tendencies in the data.

- The findings are evaluated with reference to this "best" model.

1ˢᵗ Study: Indef. neut. article: *e* vs. *es*

- There is variation in the expression of the indefinite neuter article between the forms *e* and *es* in the nominative/accusative case and after prepositions requiring accusative case.
- A linking *n* is often used between two words, one ending in a vowel and the other beginning with a vowel, to avoid a hiatus

  (1) ˌ**e**n **A**maise vs. ??**e A**maise 'an ant$_{fem.}$'

	masc.	fem.	neuter
non-oblique NOM/ACC	en Maa 'a man'	e Frau 'a woman'	e/es Chind 'a child' en/es **A**uto 'a car'
prepositional P+ACC	in en Ruum 'into a room'	in e Buude 'into a stall'	in e/es Kino 'into a cinema' in en/es **A**uto 'into a car'
prepositional P+DAT	im ene Ruum 'in a room'	in ere Buude 'in a stall'	im ene Kino 'in a cinema'

Table 1: Partial paradigm of the indefinite article in the Wil dialect of Swiss German, variation in the neuter article highlighted in red.

Indefinite neuter article: *e(n)* vs. *es*

- There are 4584 examples with the indef. neuter article. There is a clear preference for *e(n)* over *es*: 3460 *e(n)* (75.5%) – 1124 *es* (24.5%).

- The use of *e* is nearly systematic in combination with *bitzeli* 'bit$_{Dim}$', *paar* 'a few', *Wiili* 'while$_{Dim}$', Zitli 'time$_{Dim}$': n=561; 541 *e(n)* (96.4%) – 20 *es* (3.6%).

    - These "formulaic" expressions generally occur only in the NOM/ACC, and they are not modified by adjectives.

- Excluding "formulaic" expressions from the variable context: n=4023; 2919 *e(n)* (72.6%) – 1104 *es* (27.4%).

Hypotheses (i)

- Are there any factors that influence the use of the two variants?

- H1: Apparent time

  The year of birth of a speaker may influence the use of *es/e(n)* (change in progress).

- H2: Morphological/syntactic context

  The ADJ-context may influence the use of *es/e(n)*:
  indef. Det ADJ N, because ADJ is gender-marked.

  (2a)   e**s**/e luschtig**s** Chind 'a funny child'
  (2b)   en luschtig**e** Maa 'a funny man'
  (2c)   e luschtig**i** Frau 'a funny woman'

Hypotheses (ii)

- H3: Morphological context

  The DIM-context may influence the use of *es/e(n)*: indef. Det N.DIM, because all nouns in the diminutive are neuter.

  (3a)   e**s**/e Chind**li** 'a child.DIM'
  (3b)   e**s**/e Männ**li** 'a man.DIM'
 (3c)   e**s**/e Fraue**li** 'a woman.DIM'

- H4: Phonological context

  Whether or not the indef. neut. article precedes a Noun (or ADJ) with an initial *sch* or *s* may influence the use of *es/e(n)*: indef. Det N (or ADJ) +initial sibilant, because *e* may be easier to pronounce than *es*.

  (4)   e**s Sch**pilzüüg/e Schpilzüüg 'a toy'; e**s S**ignal/e Signal 'a signal'

Results: H1 - Apparent time

- There is not enough evidence to claim that the age of the speaker plays a role.

  - Adding 'Year of Birth' to the model that best captures the data distribution results in an insignificant effect.

  - The factor was significant in some models that were constructed for the data. If the effect is real, it is quite small. Only more data could detect such a small effect.
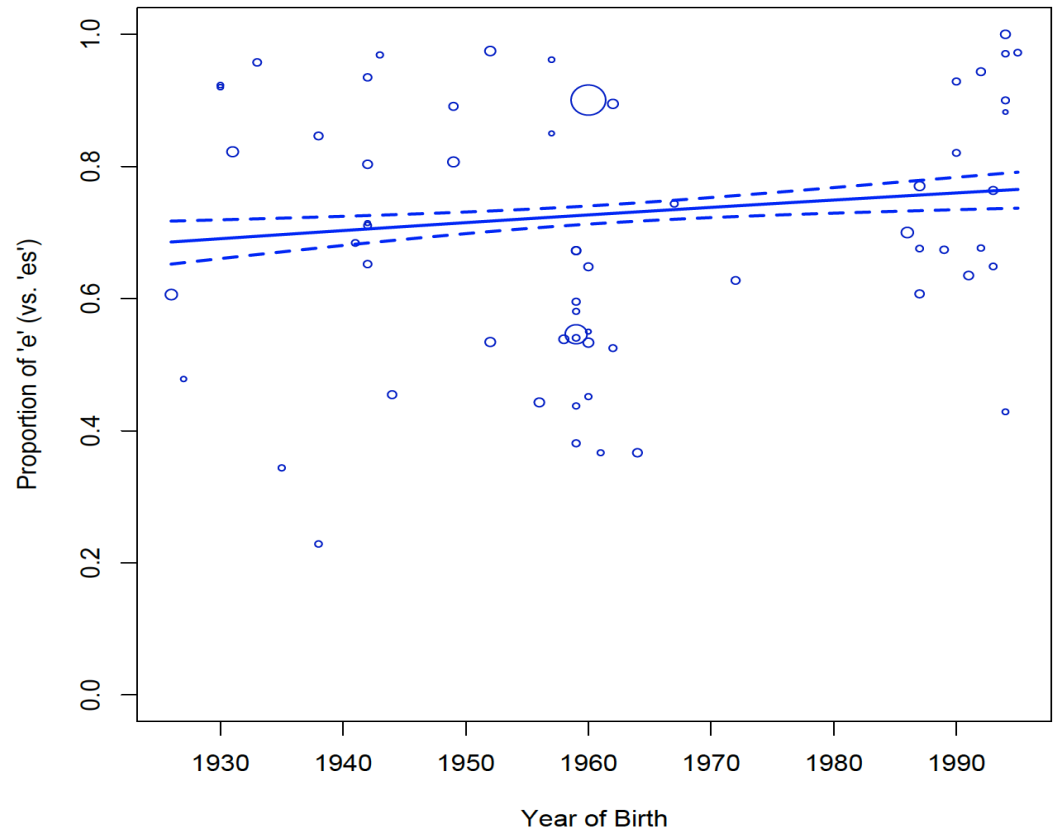
Figure 1: Logistic regression model predicting the frequency of *e(n)* (vs. *es*) from year of birth alone.

Results: H2 - ADJ-context

- The use of *e(n)* is significantly more probable when the indefinite determiner is followed by an adjective.

  - The difference was highly significant in virtually all logistic regression models that were constructed and remains significant in our best, mixed effects model with random speaker intercepts ($p < 0.001$).
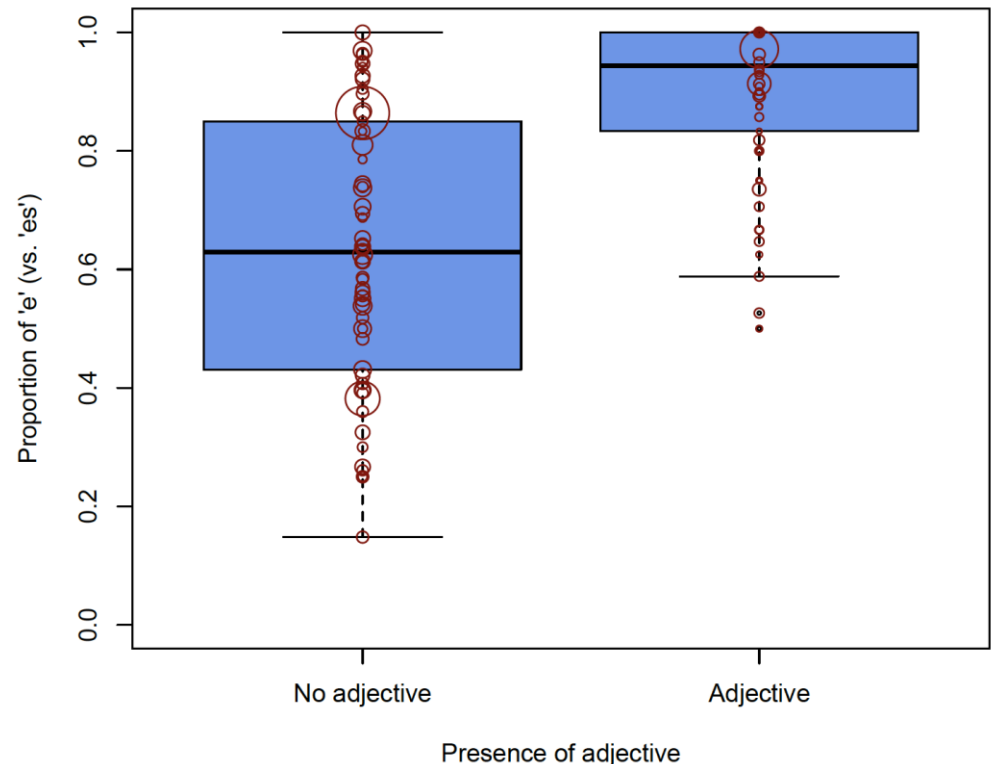

2924 examples: 1002 *es* – 1922 *e(n)*   1099 examples: 102 *es* – 997 *e(n)*

Figure 2: Boxplot for the use of *e(n)* (vs. *es*) by adjective. Red dots show the mean for individual speakers.

Results: H3 - DIM-context

- Diminutives do not significantly influence the variation between *e(n)* and *es* when added to the best model.

  - The effect is significant in some models, but loses its predictive power in particular once examples are grouped together by individual speakers. This shows that it is necessary to take random speaker effects into account.
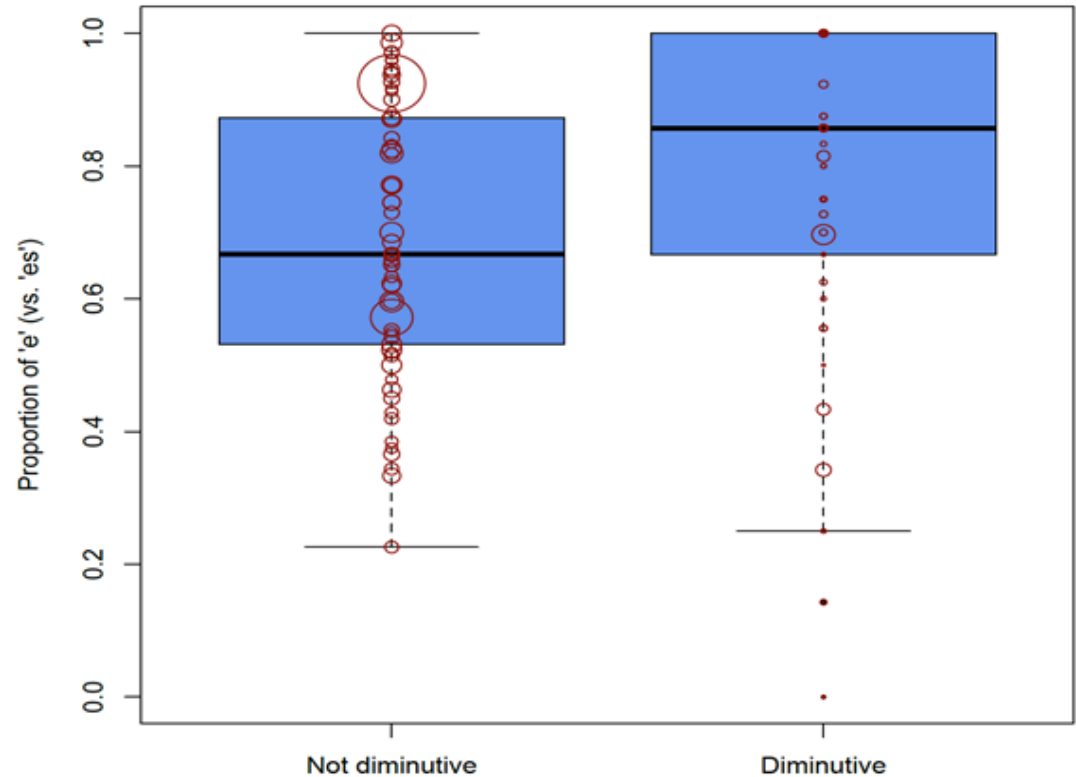

3574 examples: 975 *es* – 2599 *e(n)*

449 examples: 129 *es* – 320 *e(n)*

Figure 3: Boxplot for the use of *e(n)* (vs. *es*) by diminutive.

Results: H4 - Following sibilant

- The use of *e(n)* is significantly more common when the indefinite determiner precedes a word with an initial sibilant.

  - The difference was highly significant in virtually all logistic regression models that were constructed and remains significant in our best, mixed effects model with random speaker intercepts (*p* < 0.001).
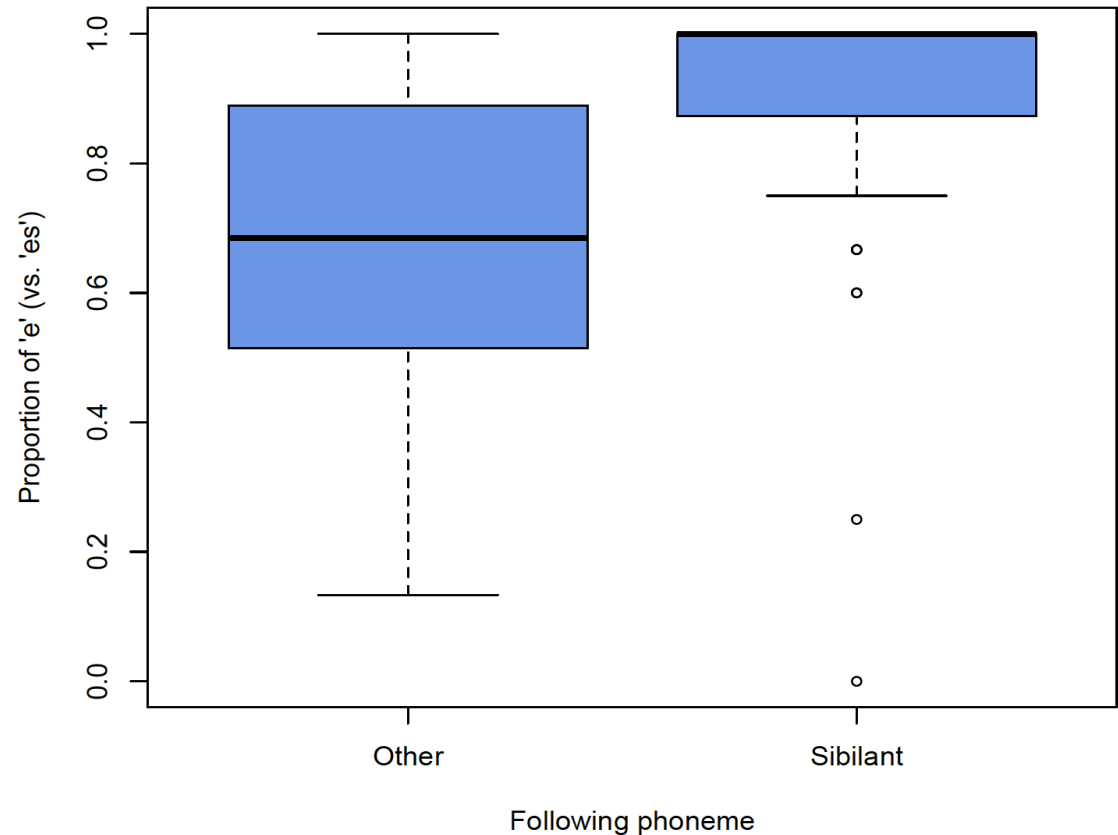
3645 examples:
1071 *es* – 2574 *e(n)*

378 examples:
33 *es* – 345 *e(n)*

Figure 4: Boxplot for the use of *e(n)* (vs. *es*) by phonological context

15

Discussion

- The effect of a speaker's age is too weak to be postulated with confidence. There is thus not enough evidence to assume a linguistic change.
  - It remains possible that the use of *e(n)* (vs. *es*) is increasing very slightly. Such a small effect could only be detected with a larger data set.

- Linguistic factors that significantly favour the use of *e(n)*:
  - The presence of an adjective after the determiner
    - Possible functional explanation: The gender-ambiguous determiner is more common because neuter morphology is already expressed on the adjective.
      - But: Morphology that marks gender unambiguously on the noun (the diminutive suffix) does not have the same significant effect.
  - A following noun or adjective starting with a sibilant

    Ease of pronunciation

2ⁿᵈ Study: 1sg of *go* 'go' (*gang* vs. *gò*)

- There is variation between *gang* and *gò* for the first person singular form of the verb *go*.

	Singular (SG)		Plural (PL)	
1ˢᵗ	(i)	gang/gò	(me)	gönd
2ⁿᵈ	(du)	gòsch	(er)	gönd
3ʳᵈ	(er/si/es)	gòòt	(si)	gönd

Table 2: Paradigm of *go* (present tense) in the Wil dialect of Swiss German, variation in the verb form highlighted in red.

1sg of *go* 'go' (*gang* vs. *gò*)

- There are 652 examples. There is a preference for *gang* over *gò*: 464 *gang* (71.2%) vs. 188 *gò* (28.8%).

- Are there any factors that influence the use of the two variants?

Hypotheses

- **H1**: Effect of Age

  The year of birth of a speaker may influence the use of *es/e(n)* (change in progress).

- **H2**: Clause Type

  The finite verb does not occupy the same position in a matrix clause and in an embedded clause. Clause type may influence the use of *gang/gò*.

(5a)   Hüt gang/gòn i i d Schtadt.   (V2: non-subject-initial)
(5b)   I gang/gò hüt i d Schtadt.   (V2: subject-initial)
(6)   I ha der dòch gsait, dass i hüt i d Schtadt gang/gò. (VE: verb-final).

- **H3**: The doubling verb *go*

  Two types of *go* can be distinguished: main verb *go* and doubling verb *go*. The type of *go* may influence the distribution of *gang/gò*?

(7a)   Etz gang/gòn i mit de ÖV. 'Nowadays I go by public transport.'
(7b)   I gang/gò bald go raise. 'I'll soon go travelling.'

Results (i)

- H1: Year of birth is significant; younger speakers use more *gò* (vs. *gang*).

  – The use of *gang* is in decline. At this rate of change, *gò* would reach a predicted probability of 99% of use with speakers born in the year 2107.

    - Year was scaled because it occurs on a large scale making it hard to estimate variances.

- H2: Clause type is significant; main clauses have an overall higher probability of *gò* (vs. *gang*) than embedded clauses.

  – There is an important interaction between Clause type and Year of birth: As embedded clauses catch up with main clauses, the use of *gò* (vs. *gang*) increases at a faster rate of change.

Results (ii)

- H3: Doubling is significant; the finite form of *go* in the doubling verb construction is overall more likely to be realized as *gò* (vs. *gang*).

- There are clear preferences among speakers to use one or the other form, necessitating the inclusion of random speaker intercepts.

  - Intraclass Correlation Coefficient (ICC): 94%, suggesting substantial variability between individuals.

Model illustration
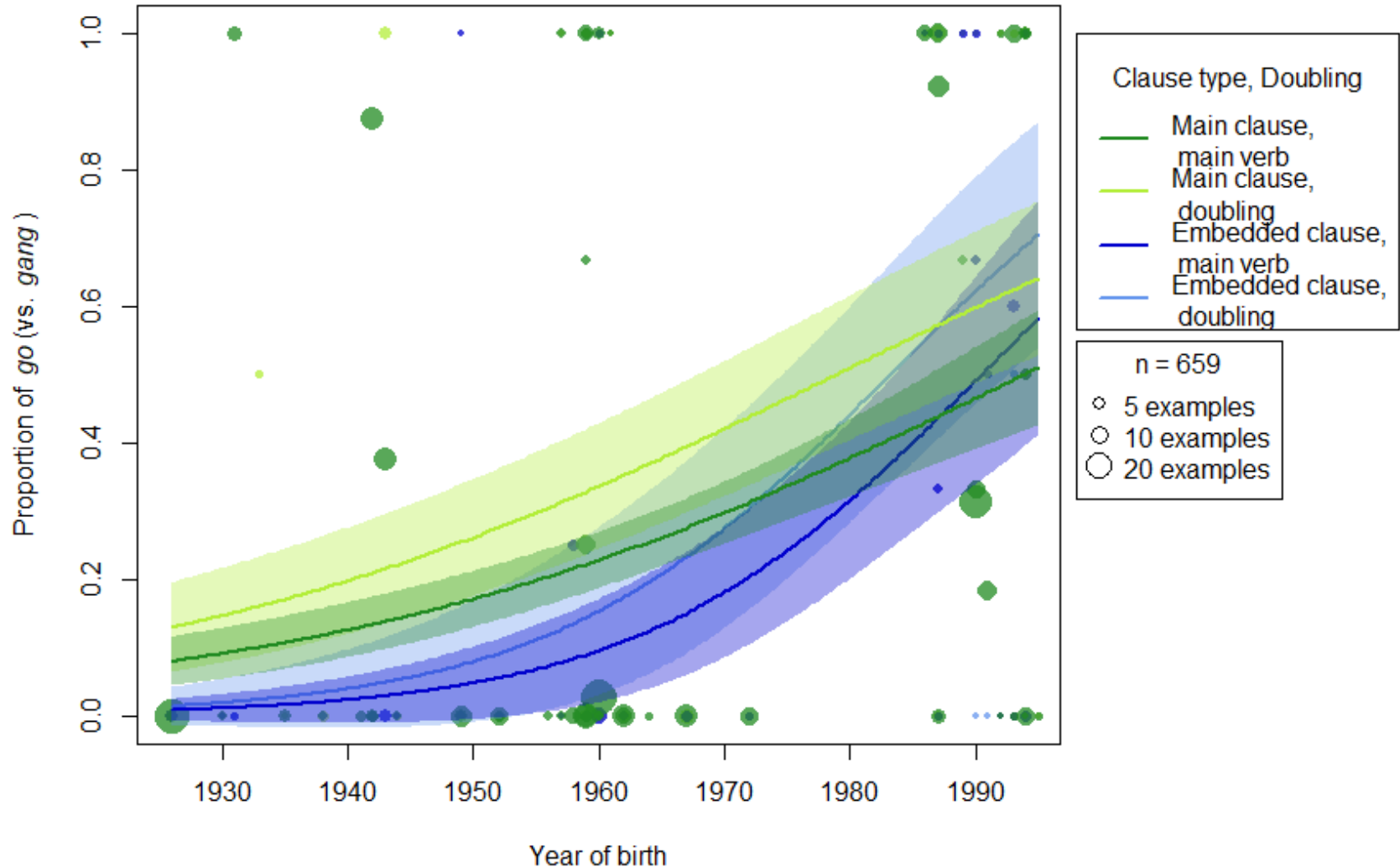
Figure 5: Logistic regression model predicting the frequency of *gò* (vs. *gang*) from year of birth, clause type and doubling.

Discussion

- The *gang/gò* variable seems to represent a change in progress, with *gò* replacing *gang*.
    - The younger the speakers are, the more they use *gò.*

- The use of the two variants shows properties that have been related to change:
    - Main clauses are innovative, subordinate clauses are conservative (cf. e.g. Bybee 2002): Higher frequency of *gò* in matrix clauses than in embedded clauses.
    - More frequently occurring items are more resistant to change than less frequently occurring items (cf. e.g. Bybee 2006): In doubling-verb contexts (n=121), the innovative form *gò* is more frequent than in main verb contexts (n=531).

Conclusions

- The nature of the two phenomena examined seems different:
  - Indefinite neuter article *e(n)/es*:
    - No clear apparent-time effect, probably no change in progress.
    - The use of the two variants is influenced by functionally motivated factors: adjacent sound, overt reduplication of morphology.
  - 1 sg *gang/gò*:
    - A clear apparent-time effect suggesting a change in progress.
    - Two other properties that would be in line with this: The innovative form is more advanced in main clauses and with the less frequently used function.
    - An open issue: Why is *gang* replaced by *gò*?
      - *Gò* is regular, *gang* is not. And: Formal similarity between the two items in the doubling-verb context.

- Next steps:
  - Additional independent variables?
  - Further areas of morphological and syntactic variation.

Merci vilmòl !

Appendix 1:

- Proposed best model for *e(n)* vs. *es*

```
Random effects:
 Groups Name          Variance Std.Dev.
 Name   (Intercept) 1.513    1.23
Number of obs: 4050, groups:  Name, 62

Fixed effects:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.4643     0.1974   12.49   <2e-16 ***
Adjnoadj     -1.8539     0.1216  -15.25   <2e-16 ***
SoundS        1.7245     0.2055    8.39   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    AIC      BIC   logLik deviance df.resid
 3807.3   3832.5  -1899.6   3799.3     4046
```

- Some model statistics:

    ICC: 31.2% (suggests some clustering within speakers)
    $R^2_{marginal}$ = 0.171, $R^2_{conditional}$ = 0.43 (model fit is ok, are important predictors missing?)
    C-index: 0.822 (acceptable predictive accuracy)

- Notes:

    This model includes only non-formulaic *e(n)/es*.
    YearBirth, Dimunitive and / or Gender added to this model will not be significant.

Appendix 2:

- Proposed best model for *go* vs. *gang*

```
Random effects:
 Groups Name            Variance Std.Dev.
 Name   (Intercept) 48.12       6.937
Number of obs: 659, groups:  Name, 58

Fixed effects:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)          -5.2491     1.9070  -2.753  0.00591 **
YearScaled            3.5509     1.1414   3.111  0.00187 **
Clausemc              1.4800     0.6557   2.257  0.02401 *
Doublingverb         -1.2969     0.4890  -2.652  0.00799 **
YearScaled:Clausemc  -1.0333     0.5258  -1.965  0.04938 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

   AIC      BIC   logLik deviance df.resid
 372.0    398.9   -180.0    360.0      653
```

- Some model statistics:

    ICC: 93.6% (suggests very substantial clustering within speakers)
    $R^2_{marginal}$ = 0.133, $R^2_{conditional}$ = 0.944 (model fit is good with random intercepts)
    C-index: 0.977 (good predictive accuracy)