Studying Gorgia Toscana: linguistic gradient phenomena as categorical, continuous, or ordinal variables?

Why am I here?

Introduction to our research(es)

(Avano, Avesani, Vayra, 2021; *in press*)

AIM: observing if moving in cities where other varieties are spoken brings speakers to level traits of their own variety due to accommodation processes (Giles and Howard, 1973).

Study objects:

1– Gorgia Toscana (GT) in 4 Florentine subjects who have studied in Bologna for 5 years compared to a control group (**students' study**)

2- Gorgia Toscana (GT) in 4 Florentine subjects who *have* been working in Bologna for more than 20 years compared to a control group (**workers' study**)

We wanted to see if people living in Bologna reduced (levelled) the GT trait or not if it was influenced by linguistics and sociolinguistics factors.

Limit (in common with other studies on accommodation): we can't really say the differences among groups are due to levelling, we can only look if the context is a predictor of the GT trait in the subjects' speech.

Gorgia Toscana

Spirantisation of voiceless stops in intervocalic position: /k/ /t/ /p/ when preceded and followed by vowels realised as follows:

- fuoco /fuoko/ [ˈfwɔːxo] [ˈfwɔːho], «fire»
- foto /fɔto/ [fɔːθo] «photo»
- capo /kapo/ [kaːφo] «head»

It can bring to many realisations, on a **weakening** (or lenition) continuum (Marotta, 2001, Sorianello 2001).

In each stdy, we had 240 intervocalic stops realised by 8 speakers, for an amount of 1920 occurrences. We classified them as follows with

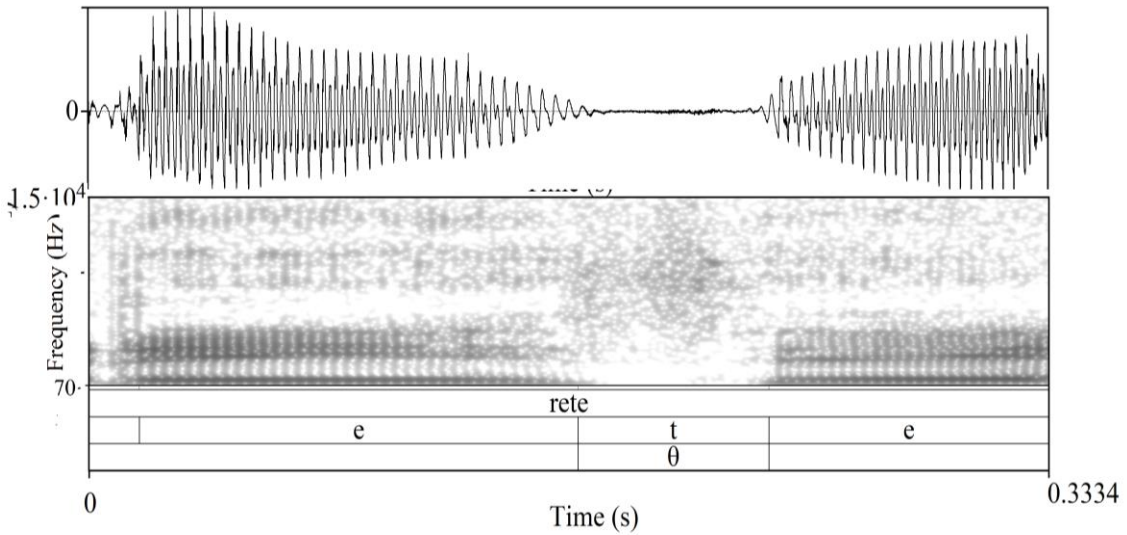
spectrographic analysis, on Praat (Boersma, P., Weenink, D. (2023):

lenition

1. Stops
2. Lenes
3. Semifricatives
4. Voiceless Fricatives
5. Voiced fricatives
6. Approximants

no GT

GT



We wanted to verify if there was a difference in lenition between subjects living in Bologna and subjects living in Forence

if this difference was influenced by other linguistics and sociolinguistics factors.

Looking for a model with the following :

Fixed factors:

1. group («fuorisede»/«controllo»
2. gender (female/male)
3. level of istruction (degree/high school diploma) – only for worker study.
4. consonant (/k/, /t/, /p/)
5. lexical stress
6. word position in the sentence

between-speaker

within-speaker

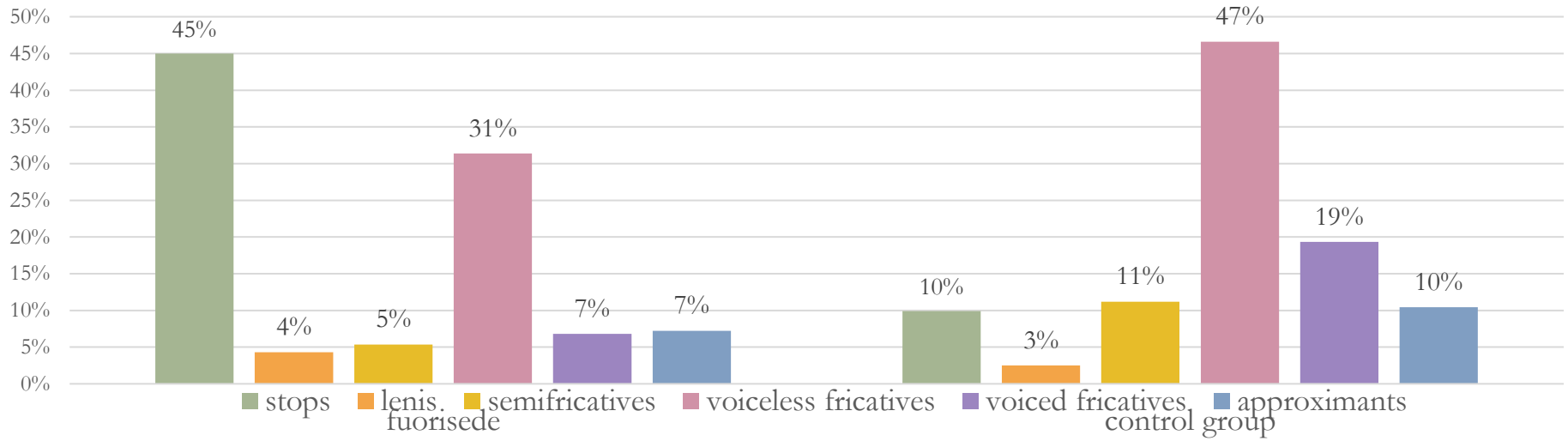
Random factors

1. Speaker
2. Word

Dependent variable? GT but how?

Student study results

	stop	lenis	semifricative	voiceless fricative	voiced fricative	approximant
fuorisede	430(45%)	41(4%)	51(5%)	300(31%)	65(7%)	69(7%)
control group	95(10%)	24(3%)	107(11%)	446(47%)	185(19%)	100(10%)
total	525 (27%)	65 (3%)	158 (8%)	746 (9%)	250 (13%)	169 (8%)



- Florentine fuorisede students realised higher percentages of stronger allophones than control group
- Students from the control group realized higher percentages of lenited allophones than florentine fuorisede students.

Which model should we use to verify if fuorisede lenited stop significantly less than control group ?

Strong difference in distribution: 65 lenes in total vs 746 voiceless fricatives

First hypotheses (July, 2021) -> on workers data

We thought using allophonic realization as a categorical variable (**multinomial regression**) wold mean to check if there is an allophonic difference on distribution, but it does not consider the allophones' gradient lenition.

Ordinal variable -> one **ordinal logistic regression model** (on jamovi, Ripley, B., 2018) to observe if there was a difference in overall lenition, but without speaker as a random factor.

but with 8 speakers, ignoring the speaker effect brought the model to overestimate the effect of interspeaker factors

Second hypothesis (february 2023) -> two different RQs depending on factors

For within-speaker factors :GT as an ordinal variable: we used **Odinal logistic model** (Ripley, B., 2018) as speaker effect is less severe than for between speaker factors and we had a strong interest on linguistic factor effect on lenition degree.

Omnibus Likelihood Ratio Tests

Predictor	χ^2	df	p
group	45.4797	1	< .001
stress	0.0582	1	0.809
place_of_articulation	114.3223	2	< .001
Prosodic_costituent	12.8728	3	0.005
group ✻ place_of_articulation	4.5475	2	0.103
group ✻ stress	0.0804	1	0.777
group ✻ prosodic_costituent	19.9436	6	0.003

For between-speaker factors GT as nominal variable : **General mixed model** (Gallucci, M., 2019) as there was a strong need to consider the effect of speaker -> observation if there was a difference in applying GT (no GT/ GT)

	X^2	df	p
group	3.881	1	0.049
genere	0.125	1	0.724
group ✻ genere	0.125	1	0.724

Third hypothesis (may 2023) -> 2 RQs

- *GT as nominal variable: Generalized mixed model* (Gallucci, M., 2019) considering all the factors -> do sociolinguistic and linguistic factor influence GT application(GT/no GT)?

Variabile	*Estimate*	*StandardError*	*exp(B)*	*z*	*p*
(Intercetta)	3.024	1.183	20.56840	2.56	0.011
Group [fuorisede – control group]	-4.895	2.364	0.00749	-2.07	0.038
phoneme [t – p]	-0.562	0.204	0.57033	-2.75	0.006
phoneme [k – p]	2.104	0.308	8.19871	6.83	< .001
group ✻ phoneme [fuorisede – control group ✻ t – p]	-0.730	0.408	0.48174	-1.79	0.073
group✻ phoneme [fuorisede – control group ✻ k – p]	-1.623	0.616	0.19725	-2.63	0.008

Group phoneme and word position as significative factors, interaction group*phoneme significative

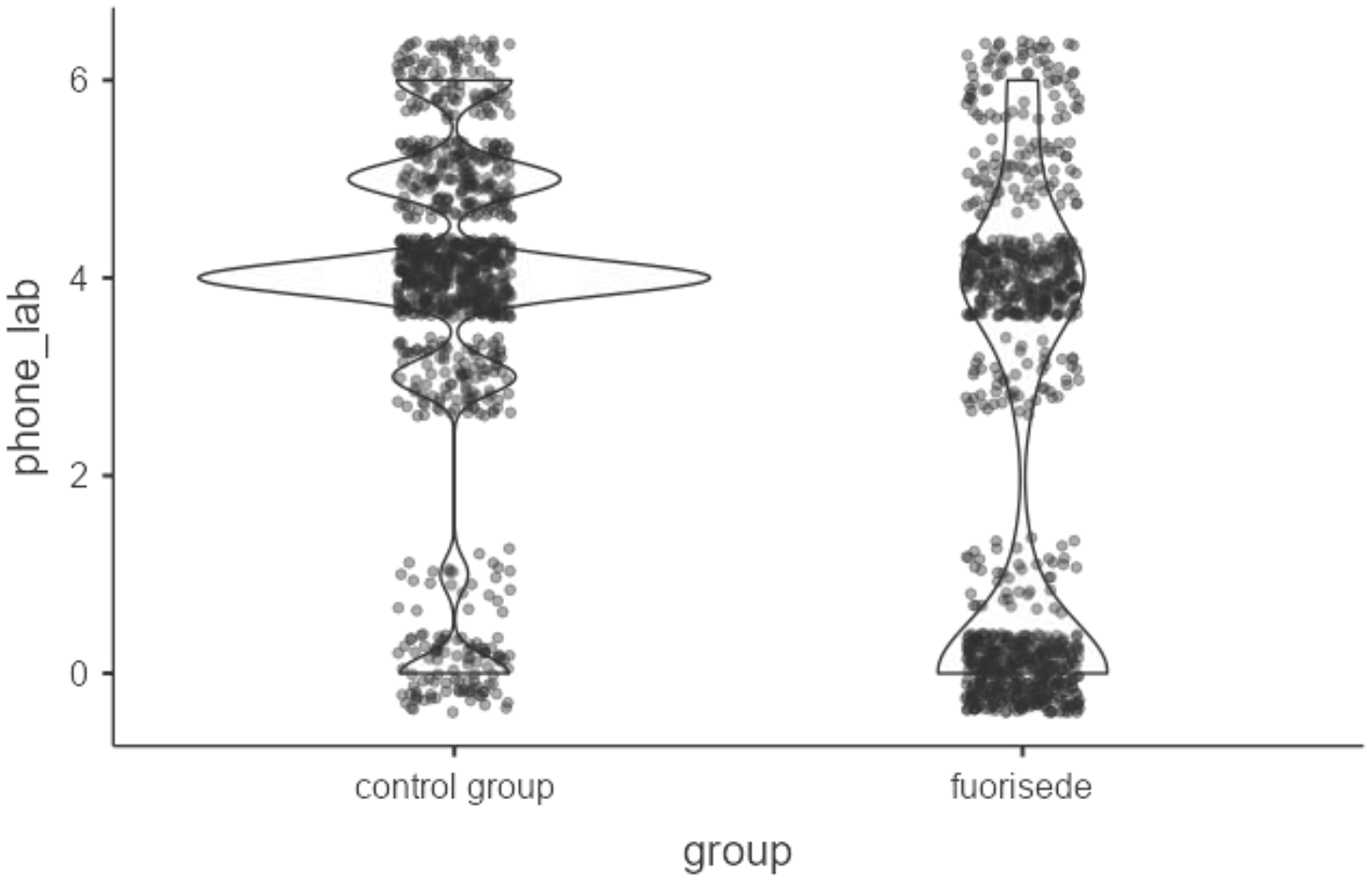
GT as continuos variable : **Mixed model** (Gallucci, M., 2019) introducing all the factors - (Bross, 2019) >do sociolinguistic and linguistic factors influence lenition?

In this case we trasformed the allophones in integer numbers:

Stops= 0, lenes =1, semifricatives =2 … approximants =5)

Variabile	*Estimate*	*StandardError*	*t*	*p*
(Intercept)	2.3348	0.4955	4.71	0.002
Posizione [*U* – *I*]	-0.1241	0.0581	-2.13	0.033
Posizione [*φ* – *I*]	0.1838	0.0582	3.16	0.002
Posizione [*0* – *I*]	-0.0698	0.0580	-1.20	0.229
Phoneme [/t/ – /p/]	0.1334	0.0502	2.66	0.008
Phoneme [/k/ – /p/]	1.0194	0.0503	20.26	< .001

Phoneme and word position as significative factors


phone_lab
6
4
2
0
control group
fuorisede
group

Ordinal mixed model (june 2023)

GT as ordinal variable (Franny D. at al., 2018) ordinal logistic regression with speaker as random factor on R (Christensen 2018)

```
model2 =clmm(phone_lab ~ group + phoneme + posizione + g
roup*posizione + group*phoneme +(1|speaker), data=accomo
damento)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
groupfuorisede	-2.44805	1.96689	-1.245	0.213268	
phonemep	-2.40530	0.16236	-14.814	< 2e-16	***
phonemet	-1.97625	0.15815	-12.496	< 2e-16	***
posizionei	0.63760	0.17604	3.622	0.000293	***
posizionep	0.88573	0.17810	4.973	6.58e-07	***
posizionez	0.24364	0.17511	1.391	0.164127	
groupfuorisede:posizionei	-0.51594	0.28833	-1.789	0.073550	.
groupfuorisede:posizionep	0.33585	0.28717	1.170	0.242197	
groupfuorisede:posizionez	-0.08708	0.28672	-0.304	0.761355	
groupfuorisede:phonemep	-0.42008	0.25165	-1.669	0.095055	.
groupfuorisede:phonemet	-0.60188	0.25166	-2.392	0.016776	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:

	Estimate	Std. Error	z value
0|1	-4.6583	1.3928	-3.345
1|3	-4.2083	1.3919	-3.023
3|4	-3.2357	1.3903	-2.327
4|5	0.6183	1.3882	0.445
5|6	2.4211	1.3894	1.743

Phoneme and position as significative factors but also group*phoneme

Assumption test (Bross 2019)

To test if the effect of the predictors […] are constant for each increase in the level of the response. (p.27)

```
> nominal_test(model2.clm)
Tests of nominal effects
formula: phone_lab ~ group + phoneme + posizione + group * posizione + group * phoneme
             Df  logLik    AIC     LRT  Pr(>Chi)
<none>           -2697.8 5427.6
group         4 -2634.5 5308.9 126.674 < 2.2e-16 ***
phoneme
posizione    12 -2666.2 5388.5  63.108 6.088e-09 ***
```

"It means that participants used the scale differently." (p.28)

```
> scale_test(model2.clm)
Tests of scale effects

formula: phone_lab ~ group + phoneme + posizione + group * posizione + group * phoneme
             Df  logLik    AIC     LRT  Pr(>Chi)
<none>           -2697.8 5427.6
group         1 -2651.4 5336.8  92.841 < 2.2e-16 ***
phoneme       2 -2669.1 5374.2  57.414 3.409e-13 ***
posizione     3 -2670.7 5379.5  54.134 1.051e-11 ***
group:posizione  7 -2625.6 5297.2 144.385 < 2.2e-16 ***
group:phoneme    5 -2629.2 5300.4 137.166 < 2.2e-16 ***
```

The model does not satisfy any assumption test…
the assumption test can be done only for the ordinal logistic regression (without the random factor)

Who should we follow?

Franny D. at al., 2018

- Apply mixed ordinal logistic regression to a linguistic trait
- Do not test the assuptions

Bross (2019)

- Apply mixed ordinal logistic models to a rating problem
- If the assumption are violated it suggests to use mixed models

Do the assumption tests matter only when we are dealing with ratings ?

or

If the assumptions aren't met the model doesn't work regardless the nature of the data?

Are ordinal models built for rating data? Is it ok to apply them on linguistic traits?

We tried them all!

We considered GT as a categorical, ordinal, nominal and ratio variable, then again ordinal…Never squaring the circle.

- Is it ok to use a mixed model as we did?
- Would it be better to use a ordinal mixed model even if it violates some of the assumptions (or are there ways ti solve it?)
- Are there other ways we did not consider?

A way we did not consider was to take phone duration as ratio dipendent variable, as it has been found to correlate with GT lenition but here too there are coarticulation problems...

Hoping for some advices,
Thank you for listening!

Riferimenti bibliografici

Boersma, P., Weenink, D. (2023). PRAAT: doing phonetics by computer [software]. Versione 6.3.03. https://www.praat.org/

Christensen, R. H. B. (2018). ordinal - Regression Models for Ordinal Data. R package version 2018.8-25. http://www.cran.r-project.org/package=ordinal/.

Gallucci, M. (2019). *GAMLj: General analyses for linear models*. [jamovi module]. Retrieved from https://gamlj.github.io/.

Marotta G. (2001-2002) , *Non solo spiranti*. *La gorgia toscana nel parlato di Pisa*, «L'Italia Dialettale», 62, pp. 27-60.

Sorianello, P. (2001). Un'analisi acustica della'gorgia'fiorentina. In L'Italia dialettale 62, 61-94..

Ripley, B., Venables W., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2018). *MASS: Support Functions and Datasets for Venables and Ripley's MASS*. [R package]. Retrieved from https://cran.r-project.org/package=MASS.

The jamovi project (2021). *jamovi*. (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org.